CS 4202 – Project Report

Dialogue Act Recognition for Text Based Sinhala



By group - Sanwada

**Palihakkara S. - 100366C**

**Sahabandu D.M. - 100466H**

**Bandara B.M.C.K. - 100052K**

**Ahsan M.S.A. - 100022U**

Project Supervisor

**Dr. Surangika Ranathunga**

# **Preface**

This report is submitted in partial fulfilment of the requirements of the research and development project “Dialogue Act Recognition for text based Sinhala” which is a part of the B.Sc. Engineering Degree in the Computer Science and Engineering, University of Moratuwa. This contains the work done from June 2014 to February 2015.

This research and development project was an awesome experience to us, and gave us a lot of opportunities to understand how the research world work and how to carry out a research based on the literature. We consider this as an extremely important part of our degree program. Our Supervisor, Dr. Surangika Ranathunga has guided and advised us throughout this journey to successfully carry out this research and development project according to the requirements of the module. We believe this has opened up our eyes to the research world and our research and development skills were significantly improved during this period.

The report contains five chapters. The first chapter presents

# **Acknowledgement**

Throughout the journey of final year project, we had to take the help and guideline of many great individuals, who deserve our greatest gratitude. First and foremost we offer our sincerest gratitude to our supervisor, Dr Surangika Ranathunga, who has supported us throughout this research and development project with her knowledge, patience and kind motivation. I attribute the level of our final year project to her encouragement and effort and without her this project too, would not have been completed.

We wish to express our heartiest gratitude to Prof. Rev. Okkampitiye Pannasara, Deaprtment of Sinhala, and University of Kelaniya and Mrs. C. Panangala, a Sinhala language teacher for providing us required Sinhala linguistic knowledge and suggestions to solve the issues we had. We make this an opportunity to appreciate our friends V.N Bandara, J.S Arugoda, S.A.W Gamage and J.P.S Bashani for tagging our corpus as volunteers with patience and support. Further, we would like to acknowledge Senuri Sahabandu for proof reading our reports.

We would like to be grateful for our alma mater, Department of Computer Science & Engineering for giving us an opportunity to do a research and development project in our university career. Also, we extend our sincere thanks to Dr. Malaka Walpola and other staff members of the department for spending their valuable time to coordination and evaluation of this research and development module in a successful manner.

# **Abstract**

This research is concerned with the classical machine learning approaches to the task of Dialogue Act Recognition for text based Sinhala. Since there is no conversational corpus exists for Sinhala, a new annotated corpus for Sinhala language is built from Sinhala subtitles for English movies. A set of dialog acts was identified based on the commonly used Dialog acts for English and linguistic characteristics of Sinhala. Feature selection was started with the common features used for English, and later on the study some exclusive Sinhala features were identified. We performed an evaluation based on features identified in utterances. Different combinations of feature set are tested on a selected classifier and some best performing features are selected based on individual feature performance. Using the best performing feature set we have performed an evaluation for accuracy and performance of the classifiers. Considering the measurements of classifier performance, the best performing classifier for Sinhala language is identified. Based on this classifier a Rest API, a desktop and a web application with functions to classify an utterance in real-time and meeting summarization are implemented.

Table of Contents

[**Preface** 2](#_Toc411284630)

[**Acknowledgement** 3](#_Toc411284631)

[**Abstract** 4](#_Toc411284632)

[1 Importance 9](#_Toc411284633)

[2 Literature Review 11](#_Toc411284634)

[2.1 Introduction 11](#_Toc411284635)

[2.1.1 Speech Acts and Illocutionary Forces 12](#_Toc411284636)

[2.2 Sinhala Language and Computing in Sinhala 13](#_Toc411284637)

[2.3 Identifying the Tag Set 15](#_Toc411284638)

[2.4 Existing Corpora 18](#_Toc411284639)

[2.4.1 VERBMOBIL 18](#_Toc411284640)

[2.4.2 SWITCHBOARD 18](#_Toc411284641)

[2.4.3 ICSI MEETING RECORDER 19](#_Toc411284642)

[2.4.4 MAPTASK 19](#_Toc411284643)

[2.4.5 Corpora for Sinhala 19](#_Toc411284644)

[2.5 Feature Selection 21](#_Toc411284645)

[2.6 Classification Techniques 24](#_Toc411284646)

[2.6.1 N-Gram Model 26](#_Toc411284647)

[2.6.2 Hidden Markov Models 27](#_Toc411284648)

[2.6.3 Naive Bayes Classifier 28](#_Toc411284649)

[2.6.4 Bayesian Network 29](#_Toc411284650)

[2.6.5 Artificial Neural Networks 29](#_Toc411284651)

[2.7 Classifier Selection 31](#_Toc411284652)

[2.7.1 Measurements 31](#_Toc411284653)

[2.7.2 Inter-annotator agreement 32](#_Toc411284654)

[2.8 Conclusion 32](#_Toc411284655)

[3 Implementation 34](#_Toc411284656)

[3.1 Work Flow 34](#_Toc411284657)

[3.2 Building Corpora for Sinhala 36](#_Toc411284658)

[3.3 Dialogue Act Tag Set Selection 38](#_Toc411284659)

[3.4 Feature Selection 50](#_Toc411284660)

[3.4.1 Identified features from related work for English 50](#_Toc411284661)

[3.4.2 Exclusive features for Sinhala 51](#_Toc411284662)

[3.4.3 Identified features 51](#_Toc411284663)

[3.4.4 Feature set selection 52](#_Toc411284664)

[4 Results 55](#_Toc411284665)

[4.1 Performance Analysis on Classifiers 61](#_Toc411284666)

[**5** **Tools** 62](#_Toc411284667)

[5.1 Desktop Application 62](#_Toc411284668)

[5.1.1 Classify utterance 63](#_Toc411284669)

[64](#_Toc411284670)

[5.1.2 Meeting summarization 64](#_Toc411284671)

[5.2 Web Application and RESTful API 67](#_Toc411284672)

[6 Future Work 69](#_Toc411284673)

[6.1 Taking Prosodic information into the context 69](#_Toc411284674)

[6.2 Identify new features 69](#_Toc411284675)

[6.3 Intelligent voice assistance for Sinhala 70](#_Toc411284676)

[6.4 Classifier optimization 70](#_Toc411284677)

[**7** **Conclusion** 71](#_Toc411284678)

[**8** **References** 72](#_Toc411284679)

Table of Figures

[**Figure 1: Language Family Diagram** 13](#_Toc411284614)

[**Figure 2: Work Flow Diagram** 35](#_Toc411284615)

[**Figure 3: Sinhala Chat Tool** 36](file:///D:\Work\Final_report_v2.docx#_Toc411284616)

[**Figure 4: Tag Set Selection Iteration 01** 42](#_Toc411284617)

[**Figure 5: Tag Set Selection Iteration 02** 45](#_Toc411284618)

[**Figure 6: Tag Set Selection Iteration 03** 47](#_Toc411284619)

[**Figure 7: Tagging tool inner view** 48](#_Toc411284620)

[**Figure 8: Tagging tool basic functionalities** 49](#_Toc411284621)

[**Figure 9** 62](file:///D:\Work\Final_report_v2.docx#_Toc411284622)

[**Figure 10** 63](#_Toc411284623)

[**Figure 11** 64](file:///D:\Work\Final_report_v2.docx#_Toc411284624)

[**Figure 12** 65](#_Toc411284625)

[**Figure 13** 66](#_Toc411284626)

[**Figure 14: Home Page** 67](#_Toc411284627)

[**Figure 15: Important Documents** 68](#_Toc411284628)

[**Figure 16: Functionality** 68](#_Toc411284629)

Table of Tables

[**Table 1: Tag Questions** 14](#_Toc411284602)

[**Table 2: Available Corpora** 18](#_Toc411284603)

[**Table 3: DA tags analysis of mostly cited research papers** 40](#_Toc411284604)

[**Table 4: Tag Set Selection Iteration 01** 41](#_Toc411284605)

[**Table 5: Tag Set Selection Iteration 02** 44](#_Toc411284606)

[**Table 6: Tag Set Selection Iteration 03** 46](#_Toc411284607)

[**Table 7: Selected Features** 50](#_Toc411284608)

[**Table 8: Cue Phrases** 51](#_Toc411284609)

[**Table 9: Individual Feature Performance** 53](#_Toc411284610)

[**Table 10: Best Performing Features** 54](#_Toc411284611)

[**Table 11: Classifier Performance** 60](#_Toc411284612)

[**Table 12: Performance analysis results for classifiers** 61](#_Toc411284613)

# Importance

Sinhala is the native language of the Sinhalese people, the largest ethnic group in Sri Lanka numbering about 16 million. Considering the other ethnic groups using Sinhala as the second Language, Sinhala can be said to be actively used by 19 million people.

Sinhala is the only language that most of the Sinhalese are fluent in. According to the Department of Census and Statistics Sri Lanka, English literacy rate of the population in 2001 is 13.63%, while the Sinhala literacy rate[[1]](#footnote-1) is 91%. Therefore there is a dire need for Sinhala language computing. With the implementation of Sinhala Unicode, the platform for this has been set. However the amount of research carried out in the area of Natural Language Processing (NLP) for Sinhala is not adequate. Unlike languages such as English, Spanish or French that are being used by larger populations in the world, Sinhala is restricted to Sri Lanka. This has an adverse impact on the progress made in Sinhala NLP research. Although there exists some preliminary-level research in areas such as Sinhala-English translation, Sinhala-Tamil (the other official language in Sri Lanka) translation, Sinhala spell checking [38], the attention paid for processing of spoken Sinhala is very low.

The aim of this study is to make use of the already existing research for Dialog Act Recognition for English and explore how it can be used in the context of Sinhala. Given the fact that dialog act recognition is an important step in understanding spontaneous dialogue, we envisage that this research would pave the path to research in areas such as meeting summarization, question-answering systems, and automated assistance.

A corpus was created from Sinhala subtitles for English movies. A set of dialog acts was identified based on the commonly used Dialog acts for English. Similarly, feature selection was started with the common features used for English, and later on the study Sinhala-specific features were identified to improve the classification accuracy. We also experimented with multiple classifiers to select the best performing classifier for Sinhala.

However, none of these are considered to be completed. When carrying out Dialog Act recognition for Sinhala, unavailability of foundational NLP research for Sinhala was a major limitation. For example, PoS tags are considered as a successful candidate in the feature set for dialog act recognition [39]. The set of PoS tags has been identified for English and there are many English PoS taggers giving very good accuracy. In contrast, Sinhala PoS tagging is at its inception stage [40]. Despite these limitations, we managed to achieve a good level of accuracy for Sinhala Dialog act recognition, by exploiting the Sinhala language-specific features. As far as we are aware, this is the first research on dialog act recognition on the family of Indo-Iranian languages. So, this research is very useful for any future researcher who carry out natural language processing related research in any cousin language of Sinhala on Indo-Iranian language branch.

# Literature Review

## Introduction

Intelligent agents have been utilized for many domains within the last decade. One of the crucial requirements of such intelligent agents (live chat based customer service providers, intelligent voice assistant agents etc.) is their ability to understand spontaneous dialogues.

To understand a spontaneous dialogue, it is important to model and automatically identify the structure of that dialogue, because it will make it easier to get a better interpretation of that spontaneous dialogue. How to model a spontaneous dialogue precisely is still an open issue, though some of the specific characteristics for modelling a spontaneous dialogue have already been identified. Among these clearly identified characteristics, “Dialogue Acts” hold an important place.

The process of identifying the Dialogue Acts (DAs) for a particular language consists of fixed set of steps [1]. That process is independent from the natural language used for the Dialogue Act Recognition.  First and foremost step of the dialogue act recognition procedure is to identify the set of DA tags that is relevant for the task. After that, relevant informative features have to be computed from the speech signal. That is a very critical step since the accuracy of identifying the Dialogue Acts heavily depend on the identified feature set. And then DA models will be trained on these identified features set.  To make the process of dialogue act recognition easier, the segmentation of the dialogues into utterances needs to be carried out independently, or alternatively realized during the recognition step with joint DA recognition and segmentation models.

Before going into the deeper level of information related to dialogue acts we first provide a brief introduction to the topic and cover some essential fundamental concepts related to dialogue acts such as illocutionary forces and speech acts to provide the background to the topic. The next section discusses existing corpora used for dialog act recognition for English. Different approaches for building a Sinhala corpus for dialogue act recognition are also discussed. The next section focuses on the dialogue act tag sets used in the process of dialogue act recognition on some of the standard corpora discussed before. It is important because almost all the research work done in the area of DA recognition considered those tag sets as a standard and used an appropriate subset of those DA tags for the research. The next section will briefly discuss the idea of inter-annotator agreement which is a useful statistic to measure the accuracy of tagging the utterances of the corpora. Following that, we will discuss the important topic of feature selection. The final section of this literature review focuses on the major standard classification techniques used for dialogue act recognition and the techniques used to rate the performance of classifiers.

### Speech Acts and Illocutionary Forces

A speech act in [linguistics](http://en.wikipedia.org/wiki/Linguistics) is an utterance that has performative function in language and communication [2]. In general, speech acts are acts of communication such as statements, requests, questions, apologies and thanking. These acts of communication are for expressing a certain attitude, and the type of speech act being performed corresponds to the type of attitude or intention being expressed. For example, a statement expresses a belief, a request expresses a desire, and an apology expresses a regret. As an act of communication, a speech act succeeds if the audience identifies, in accordance with the speaker's intention, the attitude being expressed. So dialogue acts are specialized versions of these speech acts. For example “Question” is a speech act, but “Yes-No-Question” is a dialogue act. Therefore although the number of speech acts is somewhat stable, usually ten, the number of dialogue acts depends. For example if the requirement is to process a questionnaire system, it is required to have different kinds of questions like yes-no-questions, open questions etc. However having different kinds of greetings is useless for that application. That explains how the set of dialogue acts and the size of the set depends on the application.

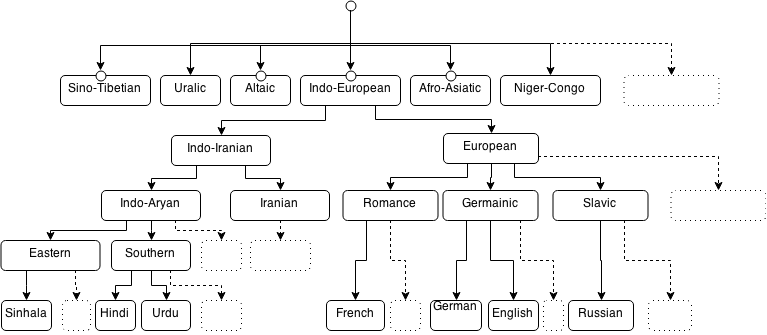
Austin [3] defines a dialogue act as the meaning of an utterance at the level of illocutionary force. The illocutionary force of an utterance is the speaker's intention in producing that utterance. Instance of a culturally defined speech act type is known as an illocutionary act, it is characterised by a particular illocutionary force. It has several types of acts, such as Asserting, Promising, Excommunicating, Exclaiming in pain, Inquiring and Ordering. For example, if we consider a speaker who asks “How is that work going on?, is it finished yet?” as a way of enquiring about the work, his or her *intent* may be in fact to make the person to finish the work. Thus the illocutionary force of the utterance is not an inquiry about the progress of the work going on, but a force for the work to be finished. A different way to define what is a dialogue act is, “a dialogue act is a specialized speech act”.

## Sinhala Language and Computing in Sinhala

Sinhala language is more than two thousand years old and it is a language akin to Hindi, Bengali and other north Indian languages. Its closest relative is the language spoken in Maldives islands, Divehi [41]. Contemporary Sinhala has been influenced by a wide variety of languages including Pali, Sanskrit, Tamil, Portuguese, Dutch and English. Sinhala alphabet is an abugida used in Sinhala writing system which is a member of Brahmic family script. It is one of the longest alphabets in use today.

Sinhala belongs to the Indo-Aryan branch of Indo-Iranian language family which along with Germanic belongs to the larger Indo-European language family. English and German languages are descendants of the Germanic branch.

(See Figure 1: Language Family Diagram)



**Figure 1: Language Family Diagram**

When considering language families, the Indo-European family, the Uralic family, the Altaic family, the Sino-Tibetan family, the Afro-Asiatic family and the Niger-Congo family can be considered as the origins of some of the major modern languages [42]. As depicted in Table 1, both Sinhala and English languages are descendants of Indo-European language family.

Following are some examples on how Sinhala differentiates from English. In English, the tag question, “isn’t it.”, “aren’t you” or “don’t they” agrees with the subject of the sentence that precedes. Its Sinhala equivalent is simply “නේ (ne)?” tagged to the end of the sentence, irrespective of its subject. Some examples are provided in Table 1.

|  |  |  |
| --- | --- | --- |
| **Sentence in Sinhala** | **Phonetic Pronunciation** | **English Meaning** |
| ඔයා තේ බොනවනේ? | oya te bonava, ne? | You drink tea, don’t you? |
| අපි තේ බොමුනේ? | api te bomu, ne? | Let’s drink tea, shall we? |
| ඔයා ඇමෙරිකන් නේ? | oya aemerikan, ne? | You’re American, aren’t you? |

**Table 1: Tag Questions**

Another example is to intensify the meaning of an adjective (such as ‘large’) English speakers add ‘very’ before it: very large. Sinhala speakers have another way of intensifying the meaning of adjectives by lengthening a vowel of the adjective itself. Thus ‘loku’ (large) can be made into ‘lokuuu’ (very large).

Although there exist many dissimilarities between Sinhala and English, it is not difficult to identify some similarities between the two languages through a much closer inspection.

If we consider the phonetic pronunciation of different words, we can observe similarities in languages of the Indo-European family. For an example the English word month pronounced in German as Monat, in Welsh as mis, in Italian as mese and in Sinhala as masaya[[2]](#footnote-2).

Moreover, the set of punctuation marks used in both Sinhala and English are identical. This could be due to the influence that Colonial English had over Sinhala.

As mentioned earlier, there exist some preliminary NLP research for Sinhala. [43] [40]

## Identifying the Tag Set

As discussed in the above sections dialogue acts are the basic building blocks for the process of spoken language understanding in human conversations. So selecting an appropriate dialogue act tag set is the crucial first step in processing conversational speech. This heavily depends on the language, culture and the context of the target application/task. Kral [1] identifies the following three requirements for selection of tag set which can be applied to any language and the any context.

1. The DA tags should be generic enough to be useful for different tasks, or at least robust to the unpredictable variability and evolution of the target application.
2. The DA tags must be specific enough to encode detailed and exploitable characteristics of the target task.
3. The DA tags must be clear and easily separable, in order to maximize the agreement between human labellers.

Although as the above three rules explain,  the selection of a tag set heavily depends on the context of the target application/task, there are some tag sets that are usually used as common base lines for most tasks. In a study, what usually happens is, first these common tag sets are studied and then target specific DA tag sets specific for the context are derived.

The natural language processing community with the guidance of the Discourse Resource Initiative [4] has designed the Dialogue Act Mark-up in Several Layers (DAMSL) tag set. With this tag set, they have targeted to provide a domain-independent universal framework for dialogue annotation. Its annotation scheme was a composition of four statistically independent (orthogonal) dimensions as follows.

1. Communicative status (defines whether the utterance is interpretable, abandoned or a self-talk.)
2. Information level (provides an abstract characterization of the content of the utterance.)
3. Forward looking functions (provides a classification in a way as actions in Searle’s [2] speech act theory.)
4. Backward looking functions (defines the relationship between the current utterance and the previous dialogue acts.)

So it is possible to combine tags from different dimensions to create new tags appropriately. After DAMSL, most of the related work used an adaptation of this tag set. The Switchboard DAMSL (SWBD-DAMSL) tag set [5] is one of the widely used adaptation. It was initially designed for the domain of telephone conversation classification.

Adapting the DA tag classes (approximately 60 tags in orthogonal dimensions) of DAMSL, SWBD-DAMSL has derived 220 different dialogue act tags to tag 205,000 utterances of the Switchboard corpus. But after clustering the rarely occurring tags (130 of the tags occurred less than 10 times) the final tag set consisted of 42 classes of DA tags.

The Meeting Recorder Dialogue Act tag set [6] is another commonly adapted tag set. It was an adaptation of the SWBD-DAMSL tag set, to classify the utterances of the ICSI (International Computer Science Institute) meeting corpus [7] that consists 72 hours of naturally occurring multi-party meetings manually tagged with DAs and adjacency pairs. The importance of this tag set comes through the characteristics of the corpus because it contains natural meetings that contain regions of high speaker overlap, affective variation, complicated interaction structures, abandoned or interrupted utterances and other interesting turn-taking and discourse-level phenomena. So this tag set can be considered as a more generalized tag set compared to the other tag sets derived upon corpora that were built on top of data collected using artificial scenarios.

Another popular dialogue act tag set is the VERMOBIL DA tag set [8]. It consists of 42 dialogue acts and 18 dialogue acts in the illocutionary level. While all the above mentioned DA tag sets are developed focussing on the English this tag set was initially designed to fulfil the requirements for the German language. So it was initially designed to facilitate German and English languages. For the determination of dialogue acts of English utterances they have used the keywords while for German utterances they have used the micro and macro structural information. The basic idea of the system is to homogeneously model preference rules by taking the information from various sources. Using that decision tree, it is possible to clarify the relationship between dialogue acts, and during the tagging process the tree is parsed from root to the leaves.

For almost all the research purposes, people have used combinations and subsets of the above mentioned tag sets. To determine the size and the composition of the appropriate dialogue act tag set, it is important to consider the size of the corpus, the context of the task etc. The natural language that was used for the study of dialogue act recognition is another important factor when selecting the tag set. Because other than the VERMOBIL tag set all the other commonly used standard tag sets are defined considering the characteristics of the English language. So it is really important to consider the language native characteristics before defining the DA tag set for the study.

If we consider the Sinhala language, although there are very few similarities between Sinhala and English we can consider it as a different language with a different set of language characteristics. For example, in English it is an easy task to categorize the utterances into the categories, commands, orders and requests, because there is significant difference between command utterances, order utterances and request utterances. For example, in a request utterance the word “please” is commonly used while in an order it is rarely used. But with the native language characteristics of Sinhala it is hard task to find the separation between these three categories. With the prosodic information it might be possible, but only with the lexical and syntactical information of the utterances, task is almost impossible. So the best thing to do is combine these three categories into a single dialogue act and use it.

## Existing Corpora

Today DA modelling techniques are widely used in the speech translation systems. That is, human-to-human communication through a machine conducting language translation. Spontaneous dialogue speech corpora are essentially important to model relevant features of spontaneous speech, such as pauses, hesitations, turn-taking behaviours, etc. and dialogue structures. There are several key corpora that have been used in most work on DA modelling. Among those few are publicly available for further studies while others are restricted. [See Table. 2]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Corpus** | **Utterance Count** | **Word Count** | **Distinct Words** | **Dialogue type** |
| SWITCHBOARD[9] | 223 606 | 1 431 725 | 21 715 | Conversational |
| VERBMOBIL[10] | 3 117 | 24 980 | 959 | Task-oriented |
| ICSI MEETING RECORDER[6] |  |  |  | Conversational |
| MAPTASK[11] | 26 621 | 152 705 | 2 502 | Task-oriented |

**Table 2: Available Corpora**

### VERBMOBIL

It is a German research project, project that aims at translating the spontaneously spoken dialogues robustly and bi-directionally for German/English and German/Japanese. It produced a corpus of 168 English annotated task-oriented dialogues. In order to tag the VERBMOBIL corpus they used a total of 46 tags, which are then further clustered into 26 top-level tags.

### SWITCHBOARD

This corpus comprises 1155 annotated telephone conversations that have greater variability of topics. Due to that, SWITCHBOARD corpus exhibits greater semantic variability than any other corpus created that time. Therefore it has been a more difficult problem for accurate DA modelling. This corpus was initially tagged with 220 tags. 130 of those tags that occurred less than 10 times have been clustered. Finally that lead to the 42 larger tag classes.

### ICSI MEETING RECORDER

This project had a corpus of over 180,000 hand annotated dialog act tags and accompanying adjacency pair annotations for roughly 72 hours of speech from 75 naturally-occurring meetings. For tagging the Meeting Recorder corpus used 65 tags. Most of tags they adopted from the SWBD-DAMSL since it fits with the corpus.

### MAPTASK

Design of this corpus mainly focused on allowing the investigation of a range of issues relevant to both psychological models of human language production and comprehension and speech technology, especially as the focus on effort switched to more natural, unconstrained speech.  MAPTASK corpus runs about 18 hours of speech, which have generated 152 705 word counts. Word lists containing all the feature names were also elicited from all speakers, along with a number of 'accent diagnosis' utterances. That contains 128 task-oriented dialogues and it uses 12 distinct DA tags.

### Corpora for Sinhala

“Sinhala” is the native language for 80% of Sri Lankan citizens. In order to perform linguistic research on Sinhala language, the main pitfall is the lack of a standardized corpora for Sinhala. To create a corpora for Sinhala language one can use several techniques to collect data such as,

* Create a Sinhala chat tool and  collect chat data
* Create a tool to extract data from Sinhala News Paper / Novels etc.
* Create a tool to extract Sinhala Subtitles for foreign movies/ TV series.
* Collect telephone conversation data

Among above points, the last point is a very difficult task to perform, because it might require lot of tools to capture a high quality data set. As well as speech-to-text conversion also needed. But for Sinhala no robust method is there to perform such task. But the initial 3 points look a bit easier compared to the last technique. Creating a chat tool and collecting those data would be bit time consuming. But other remaining two facts are comparably easy to other methodologies. But finding Novels and Newspapers will be bit difficult compared to finding subtitles files on the internet, because there are huge archives existing for Sinhala subtitles for foreign movies and TV shows. In order to get those files from such archives one can create a small web crawler and then create a small program in any programming language to extract and store those sentences in a database easily.

## Feature Selection

Akker and Schulz [12] identify features as the input given for the classifier, as a vector for each word in the utterance. Features can be extracted from the word itself, timing and the prosodic information. They have further identified 4 major categories of features as they present in an utterance.

1. Time Related Features
2. Word Related Features
3. Prosodic Features
4. Online Features

Time related features are derived from the start and end times of the utterances in the corpus. ***Pause between two words*** is one feature. It is the difference between the start time of the current word and the end time of the last word by the same speaker.  ***Duration of the word***, ***Mean duration of the word*** and ***Relative duration*** of the word are other word duration oriented features which are extracted from the corpus. For deriving these features from the corpus, the utterance itself needs to have information about the speaker, the duration of each word, etc.

Words can act as features themselves. These are categorized under the word related features. ***Current word***, ***next word*** and ***previous word*** are 3 most commonly used features. Since most classifiers cannot deal with strings directly, most of the time the feature of the string is converted into a nominal feature for each word. Akker and Schulz [12] have only used words in the corpus that occurred more than 100 times as nominal feature. Using the same procedure next word and previous word features can be derived from the same utterance.

Stegeman and Akker [13] describe different perspectives of the above 3 features, considering the part-of-Speech (PoS) that the current, next and previous words are located in. PoS (Part-of -Speech) is a linguistic category of words which is generally defined by thesyntactic ormorphological behaviour of the lexical item in question. Commonly *Verbs, Nouns, Adjectives, Adverbs WH-Questions* are used as this categories. Feature ***PoS Current word*** is derived by a tagger from the current word and surrounding 6 words. The tagger used Penn Treebank English tag set containing 37 tags [14]. ***PoS previous word*** feature derived from the words preceding the current word while ***PoS next word*** feature from the words that are following the current word.

Since the 37 sized tag sets used for PoS features are too fine grained for DA tagging purpose, the Penn Treebank tag set is mapped to a 6 tag set: *Verbs, Nouns, Adjectives, Adverbs WH-Questions and Other.* Using above 6 tags, they have redefined the PoS features as ***PoS Reduced Current word, PoS Reduced Next word*** and ***PoS Reduced Previous word***. Some words in an utterance have more impact on the DA than other words in the same utterance. So as an extension to this reduction, ***PoS with Keywords*** have been introduced where some words in the corpus get their own tag.

Other than the aforementioned features, two other word related features can be identified as ***Word Repeat*** and ***Word Repeat 2,*** where the former is evaluated to *true* if the next word is same as the current word and the latter is evaluated to *true* if the previous word is same as the current word.

Prosodic Features are derived from the prosodic information stored in the utterances [15]. Word pitch and Energy information can be used as features after amplitude values of the signal have been normalized to the microphones [16]. ***Pitch Features*** and ***Energy Features*** can be evaluated using the *minimum, maximum* and the *mean* values of pitch and energy. ***Speech-flow Past, Speech-flow Future*** and ***Speech-flow Change*** featuresdefine the talking speed with respect to the other words in the surrounding phrases.

Apart from above features, Akker and Schulz [12] discussed 4 other features related to segmentation of the utterances. ***Number of words in previous segment*** feature is self-explanatory. ***Distance to the last segment*** feature is the number of words from the end point of the last segment. ***Relative position of word inside the segment*** and ***Time interval of current word to last segment*** are other segmentation related features.

In the set of experiments done by Akker and Schulz [12], they have identified the best performing feature set including *Pause between words, Mean duration of words, Specific current words, Previous words, Part of Speech information, Minimum and Mean of Energy, Speed-flow change* and *Length of a segment.*

Apart from above discussed commonly used feature sets, other researchers have done several experiments with different types of features as well. Rosset and Lamel [17] used a feature-vector consisting of *Speaker Identity, Number of utterances* and *First two words.* Lendvai [18] opted for not using DA tag of previous utterance as a feature for the current utterance, as it could introduce a cumulative error. *Utterance type, Presence/absence of Wh-Question* and *Subject type* were used as features by *Andernach* [19]. He also used two interesting features *1st verb type* and *2nd verb type* because of their potential of informatively on kinds of agents and actions. Similar use of above mentioned two features can be seen in [15] as *grammar patterns.*

## Classification Techniques

Dialog acts (DAs) represent the functional building blocks of conversations and the classification of dialog acts corresponding to assigning DA types to the individual utterances[20].How these different DA types are defined depends on many factors such as goal of the application, size of the corpus and the experimental setup. Dialogue act classification is a special case of text classification where the text to be classified is the user utterance. The state-of-art in dialogue act classification is to use all available information sources from multiple perspectives [21], including:

1. Linguistic information that can be derived from the surface form of an utterance: lexical and collocation information.

Linguistic DA classification is based on the observation that different DAs use distinctive word strings. It is known that certain cue words and phrases [22] can serve as explicit indicators of discourse structure. Similarly, we find distinctive correlations between certain phrases and DA types. For example, 92.4% of the uh-huh’s occur in BACKCHANNELS, and 88.4% of the trigrams “<start>do you” occur in YES-NO-QUESTIONS.

1. perceptual information from multiple channels available to dialogue participants, including acoustic and prosodic properties of utterances as well as information from visual and other modalities

Prosodic and acoustic DA classification is based on number of factors such as utterance duration, pitch, pauses, energy, speaking rate and gender that are computed automatically from the speech signal. Prosodic information is vital for DA classification, because word-based classification suffers from recognition errors and some utterances are inherently ambiguous based on words alone. For example, some Yes-No-Questions have word sequences identical to those of statements, but can often be distinguished by their prosodic information [23].

Contextual information obtained from the preceding dialogue context and dialogue structure, as well as global context properties like dialogue setting, knowledge about dialogue participants, and domain knowledge

Structural information like location of an utterance and the context of the utterance can be a strong predictor of the dialogue act. Based on the nature of the application, it is observed that the utterance position in a sentence as well as in a turn plays an important role when identifying its dialogue act. For example, an utterance such as “Hello” will occur at the beginning of a dialogue while an utterance such as “Have a nice day” will typically appear at the end. So, the position of utterances in a turn can also help identify the dialogue act; i.e. when there are several utterances in a turn, utterances are related to each other, and thus examining the previous utterances in the same turn can help correctly predict the target utterance. For example, the greeting “Welcome” and question “How may I help you?” could occur in the same turn.

A wide variety of machine-learning techniques have been used for DA classification tasks with various instantiations of feature-sets and target class encodings, and for dialogue processing, it is still an open issue which techniques are the most suitable for which task [25].For example, techniques based on n-gram language modelling were applied by Reithinger [25] to the Verbmobil corpus, with a reported tagging accuracy of 74.7%. Hidden Markov Models (HMM) have been tried for dialogue act classification in the SWITCHBOARD corpus by Stolcke [23], achieving a tagging accuracy of 71% on word transcripts. Another approach that has been applied to dialogue act recognition, by Samuel [26], uses transformation-based learning. They achieved an average tagging accuracy of 75.12% for the Verbmobil corpus. Keizer [27] used Bayesian Networks, applying a slightly modified version of DAMSL with an accuracy of 88% for backward-looking functions and 73% for forward-looking functions in the SCHISMA corpus. Lendvai [28] adopted a memory-based approach, based on the k-nearest-neighbour algorithm, and reported a tagging accuracy of 73.8% for the OVIS data.

The following are some of the popular classification techniques used in the literature.

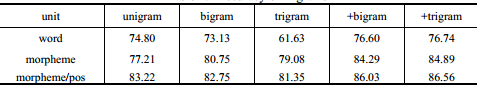
### N-Gram Model

One of the simplest techniques used in DA classification is the n-gram. N-gram is a contiguous sequence of n items from a given sequence of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application- Wiki. N-gram models can be imagined as placing a small window over a sentence or a text, in which only n words are visible at the same time. The simplest n-gram model is therefore a so-called unigram model. This is a model in which we only look at one word at a time. The sentence “This is our literature review”, for instance, contains five unigrams: “This”, “is”, “our”, “literature”, and “review”. Of course, this is not very informative, as these are just the words that form the sentence. In fact, N-grams start to become interesting when n is two (a bigram) or greater. In similar fashion, a bigram can be thought of as a window that shows two words at a time such as {(This, is), (is, our), (our literature), (literature, review)}.

In the context of DA classification, N-Gram model is based on the assumption that the current dialogue act is explicitly determined by k preceding dialogue acts [29]. Therefore the candidate for the n-th dialogue act is chosen by the principle

Capture.PNG

Here the conditional probabilities are extracted from tagged corpora by counting all existing DA sequences [34]. These are simply the number of occurrences of the sequence (cn−k+1, . . . , cn) in the training corpus, divided by the number of occurrences of a shorter sequence, (cn−k+1, . . . ,cn−1). The typical values for k would be 2(bigram) or 3(trigram). Using larger k values only make sense when longer dependencies are known to exist. The unigram models lack the disambiguation information while the trigram models seriously suffer from the data sparseness problem. As represented in the below table [30], the combination models such as the +bigram models and the +trigram models achieve much better performance than other models; because the combination models can alleviate the sparse data problem by using the unigram features, and utilize the enriched features by using the bigram features or the trigram features.



### Hidden Markov Models

Hidden Markov Models (HMMs) have remained as one of the dominant statistical modelling techniques used in modern Dialog Act recognition systems. HMM is a tool for representing probability over sequences of observations where a process is modelled as two parallel sequences of states, out of which one is observable and the other one is hidden. In this model, the state is not directly visible, but the output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore the sequence of tokens generated by an HMM give some information about the sequence of states. The Hidden Markov Model is based on following the two important assumptions related to the nature of the dialog Act [31].

* Limited Horizon Assumption - The conditional probability distribution of future states of a process depends only upon the present state, not on the sequence of events that preceded it.

markov1.PNG

* Time Invariant Assumption - Transition and generation probabilities remain stationary throughout the process.

first part.PNGtime.PNG

Here X = (X1 ,..., XT) is a sequence of random variables taking values in some finite set S =  {s1, . . . , sN), the state space.

There are two main approaches of HMM widely used in practise known as Forward-backward algorithm and Viterbi algorithms[32].Forward-Backward is used if one wants to predict the most likely token at a given time. It takes every possible sequence into account and averages over them to find the most likely token at that time. Therefore the returned sequence in not a true sequence, but a collection of the most probable tokens considering all of the possible sequences. Viterbi is used to find the most likely sequence of events. This algorithms looks at every sequence and simply selects the sequence that is most likely to be happen.

### Naive Bayes Classifier

The Naive Bayes Classifier is a simple probabilistic classifier based on the Bayesian theorem with the assumption of a high degree of independence between features (strong independence assumption). The naive Bayes classifier assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class variable (DA tag). For example, a dialog act may be considered a back channel question if it is a backchannel and a question. This classifier considers each of these features to contribute independently to the probability that this DA is a backchannel question, regardless of the presence or absence of the other features. From Bayes theorem, we have

bayes.PNG

Here fi belongs to features and C belongs to the classes (DA tag). So the most probable class can be written as

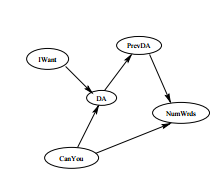
bayes1.PNG

According to the strong independence assumption, the probability of observing each feature for a given class is just the product of the probabilities of the individual features given in the class. So, the general Naive Bayes classifier can be written as [31]

naives.PNG

### Bayesian Network

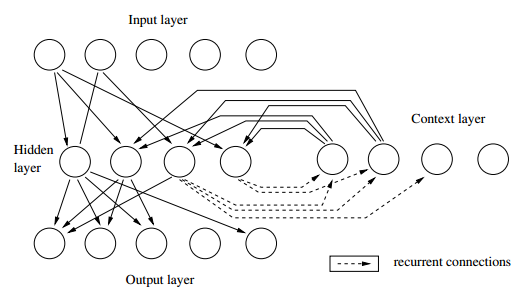
Bayesian networks or belief networks are an advancement over the naive Bayes classifier. Instead of assuming the features to be independent and the output to depend on the features, in this model the dependencies are specified by a directed acyclic graph (the dependency network). The output is regarded as an unknown feature that can both depend on and influence other features. This probabilistic graphical model is used to represent knowledge about an uncertain domain that encodes probabilistic relationships among variables of interest. Each node in the graph represents a random variable, while edges between nodes represent the probabilistic dependencies among the corresponding random variables. In a nutshell, the Bayesian probability of an event is the degree of belief in that event. The following diagram clearly illustrates the above mentioned scenario.



.

### Artificial Neural Networks

Artificial Neural Network is one of the most frequently used neural network techniques, in general as well as in DA recognition. An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the multilayer structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. The input propagates through the network layer-by-layer. The following figure demonstrates this novel architecture of a simple neural network with one hidden layer and one context layer [32].



An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons [37]. Here it is called input weights and it is adjusted dynamically according to a learning rate until it reaches close enough to the target output. Simply, when a neural network is initially presented with a pattern, it makes a random 'guess' as to what it might be. It then sees how far its answer was from the actual one and makes an appropriate adjustment to its connection weights. ANNs, like people, learn by example. It cannot be programmed to perform a specific task. The training examples must be selected carefully otherwise useful time is wasted or even worse, the network might function incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable.

## Classifier Selection

In order to select a suitable classifier to perform the classification, we need to get some measurements about how different classifiers perform on a set of utterances. A set of classifiers have to be trained and tested on a set of utterances. But again the test results affected by the feature set selection, number of training instances and classifier parameters. So it is convenient to use a fixed set of utterances as training set with selected fixed feature sets and without any classifier parameters.

### Measurements

#### Precision and Recall

In pattern recognition and data mining with classification, there are two significant parameters to evaluate the accuracy of the classification. The **Precision (**positive predictive value**)** and **Recall (**sensitivity**).** Precision is defined as the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved [33].  Precision can be seen as a measure of exactness or quality whereas recall is a measure of completeness or quantity. From the statistical viewpoint, when the scenario *all and only the relevant items are retrieved* istaken as the null hypothesis, absence of type I and type II errors corresponds respectively to maximum precision (no false positive) and maximum recall (no false negative).

*Definitions*

#### F-measure

F-measure combines the two measurements precision and recall and is taken as the harmonic mean of precision and recall [33]. F-measure reaches its best value at 1 and the worst score is 0.

The F-measure has been widely used in natural language processing literature. We can use F-measure to rank the classifiers against above mentioned feature sets and parameters.

### Inter-annotator agreement

After selecting the appropriate corpus and a tag set, the next critical step of the study is to tag the utterances of the corpus using the selected DA tag set. The tagging process should be done manually using real, non-biased people and for that, it is possible to use a single person or several people. It is really important to tag the corpus accurately in order to achieve best possible results in the final classifications. So to improve the accuracy of tagging, it is better to use several people for the task instead of a single person. So the better procedure is for several people to separately tag same part of the corpus manually and then review the results. It is possible to find an utterance where everyone used the same DA tag for an utterance and it is also possible to find the utterance is tagged using different tags by different users due to the ambiguities of the utterance or DA tag set.

So after tagging the corpus it is essential to find the comparison between the resultant tagged corpora of separate users. To evaluate this inter-annotator agreement value, Cohen [34] defined a statistic (Kappa value). The kappa value gives a percentage value of agreement between two annotators. But this method cannot be used to calculate the inter-annotator agreement between more than two users. To overcome that limitation Fleiss [35] defined a new statistic (Fleiss Kappa). That method is capable of finding the inter-annotator agreement between any numbers of annotators.

## Conclusion

In this review we have presented an extensive survey of Dialogue Act Recognition by data mining approaches and a brief review of related studies, research and techniques. We have explained the importance of recognizing Dialogue Acts in a conversation and the major steps involved with the process of identification. Extensive description about corpora used in various studies for different languages including English, German and French. We have compared the corpora available in the context against utterance count, word count and dialogue type. In this literature we pointed out that there are no corpus available for Sinhala language and we provided the approaches can be used to build a Sinhala corpus.

Selecting features for the classifiers is a very important step in the process of dialogue act recognition using classifiers. We have summarized the feature selection criteria used by the studies done in the context. Then we discussed about different classification techniques can be used and we end the review with a discussion on selecting a classifier and the parameters we can use for measure classifier performance.

# Implementation

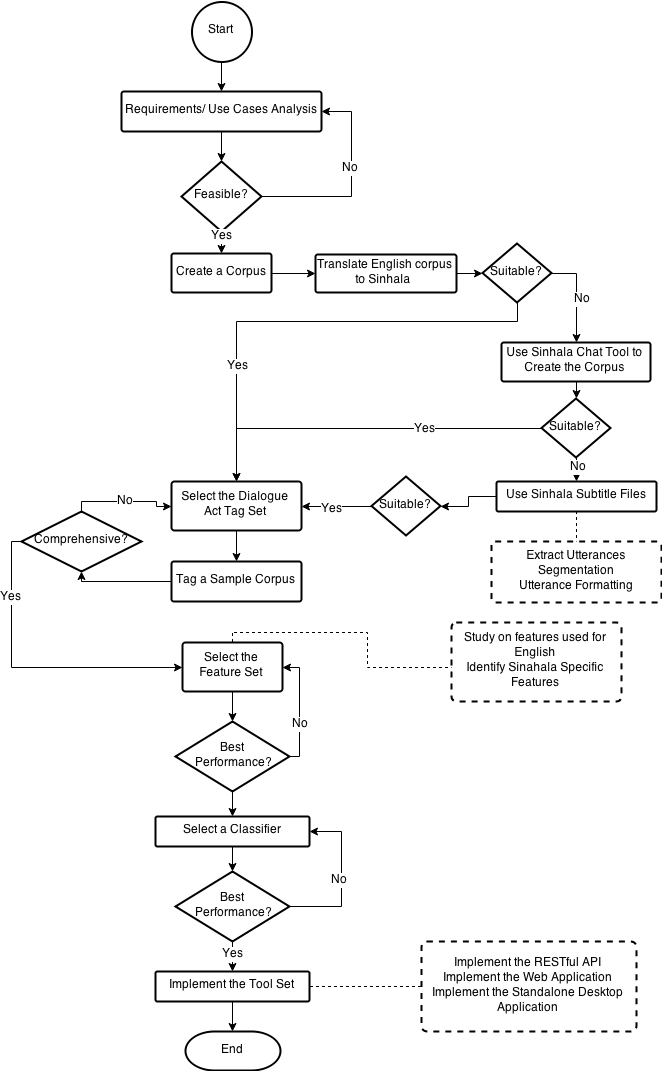
## Work Flow

For this project there were seven major milestones to achieve.

1. Corpus Building
2. Dialogue Act Tag Set Selection
3. Feature Selection
4. Classifier Selection
5. Results Analysis
6. Tool Implementation

The following work flow diagram (figure 02) depicts how these tasks were achieved. Before going into the research and development work it was required to do a requirement analysis and use case study to understand the importance and the feasibility of the project. Above section (1) on ***Importance*** of the project has already discussed the relevant details. So under this section the discussion will be started from the point of corpus building.

Under this section all the aspect of the following flow diagram will be discussed considering the methodologies that has been followed to achieve the above mentioned major milestones of the project.

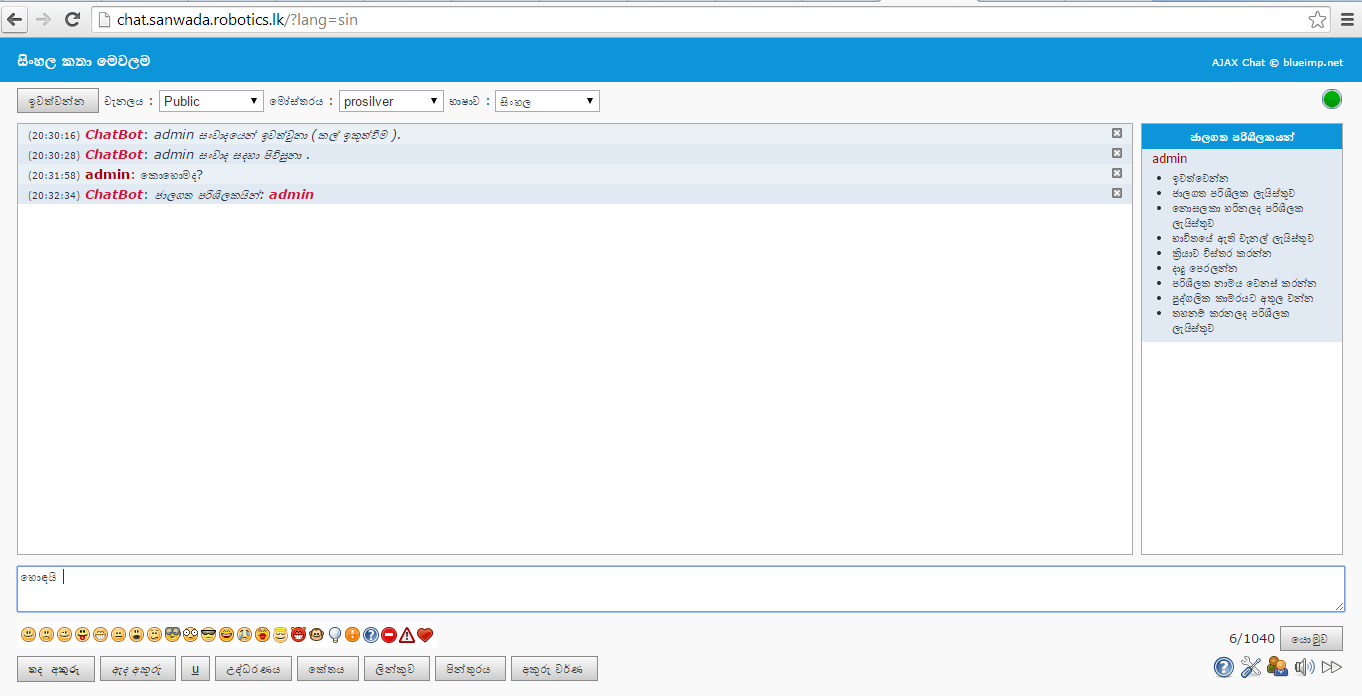


**Figure 2: Work Flow Diagram**

## Building Corpora for Sinhala

Since no corpus was available for dialog act recognition for Sinhala, it was required to build a standard corpus from the scratch. We tried out three approaches for this task.

1. Translate an existing standard English corpus
2. Sinhala chat tool
3. Sinhala movie subtitles

Finding translators was not possible so we abandoned the first option. Then we deployed a Sinhala chat tool (figure 3) for public use and collected conversations. At the beginning this approach seemed promising but the process was slow because it was difficult to get volunteers and the volunteers were tend use to use English words in the middle of Sinhala utterances. Also they used slangs and urban words more often which makes the classification more complex. Although we understand that a dialog act recognition system should accept the existence of such non-standard words, this was considered out of scope for the current research.

**Figure 3: Sinhala Chat Tool**

Then we tried to extract utterances from Sinhala subtitles of English movies. The translation of English movies is a result of a community-based crowed sourcing effort. About 10 full-time translators are contributing to this under the trade name of “baiscopelk”[[3]](#footnote-3). In Sri Lanka, there is a large population that enjoy Hollywood movies and TV series. However, their low English literacy is a problem when understanding these movies and TV series. The aim of baiscope.lk is to provide Sinhala subtitles. The subtitle creation process is governed by a set of rules and regulations. The subtitles are almost in grammatically correct Sinhala.

One issue with this method is that some movies have frequent scene changes. This is problematic for extracting consistent conversations. To overcome this we had to manually select the movies that contained long consistent scenes. We collected about 1.8 million utterances using this method for 2306 movies.

Extraction and segmentation of utterances were done manually to build a more conversation-oriented corpus. Extracting the utterances from a subtitle file consist of several steps. First step is to omit the time-related information mentioned alongside utterances. Then the filtering out of advertisements and symbolic characters takes place. Finally eliminating the improperly used punctuation marks. Such as using multiple exclamation/question marks instead of using one right after an utterance in order to emphasize the emotion conveyed in the movie scene. Segmentation is done manually by checking each line for one statement broken into few lines in the subtitles. It is a result of a scene change in the middle of an utterance in the movie. If any such lines found, can combine them into a single line.

The final corpus contains 1.8 million utterances including tagged 12,000 utterances. It is publicly available[[4]](#footnote-4) under the name of “Sanwada”. In Sinhala, the term “Sanwada” means conversation.

## Dialogue Act Tag Set Selection

For almost all the research purposes, people have used combinations and subsets of the mentioned tag sets in above section (2.3) ***Identifying the Tag Set***. There are several major factors governing the size and the composition of dialogue act tag set for the purpose. For an example it is important to consider the factors such as size of the corpus, the context of the task. The natural language that was used for the study of dialogue act recognition is another important factor. Because other than the VERMOBIL tag set all the other commonly used standard tag sets are defined considering the characteristics of the English language. So it is really important to consider the language native characteristics before defining the DA tag set for the study.

In this study to select the appropriate dialogue act tag set an iterative method was used. The process was to select an appropriate tag set for the domain and then practically try to use the tag set to determine the comprehensiveness of the selection. The process was conducted in 3 major iterations. In each iteration a sample of 500 utterances from the corpus was tagged manually using the selected dialogue act tag set. The tagging process was completed by four independent volunteer annotators

That initial tag set was filtered out considering the frequency of their presence in the related work for English. The following table 3 depicts the major dialogue acts and their presence in mostly cited related work mentioned in the following list.

1. Dialogue act modeling for automatic tagging and recognition of conversational speech [ Andreas Stockle-2000]
2. Classifying Dialogue Acts in One-on-One Live Chats[Su Nam Kim-2000]
3. Dialogue Act Classification Based on Intra-Utterance Features[Nick Webb-2000] - SWITCHBOARD Dialogue Acts(42 tag Classes)
4. Automatic Instant Messaging Dialog using statistical model and dialog acts- [Edward Ivonic-2008]
5. Text Based Dialogue Act Classification for Multiparty Meetings. [Matthias Shirberg - 2006]
6. Lexical and Discourse Analysis of Online Chat Dialog - [Eric N. Forsyth and Craig H. Martell- 2007 ]
7. Combinations of Classifiers for Automatic Recognition of Dialogue acts[ Pavel Kral-2005]
8. Lexical, Prosodic, and Syntactic Cues for Dialogue Acts [ Daniel Jurafsky - 1997]

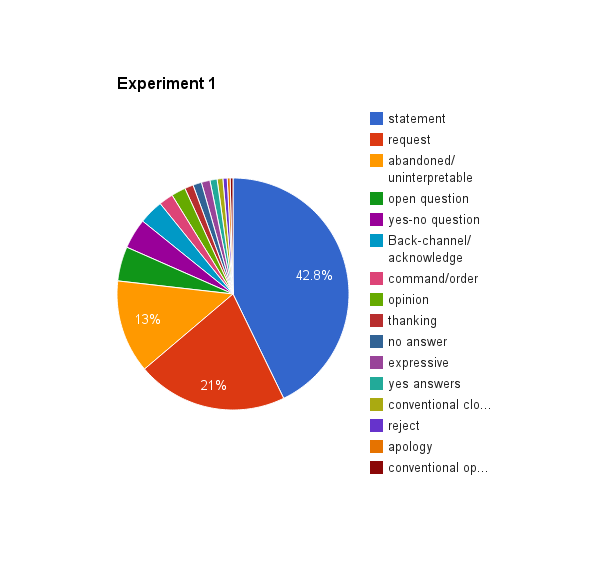
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tags** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **Sum** |
| Statement | X | X | X | X | X | X | X | X | 8 |
| Backchannel/ Ack | X | X | X | X | X |  |  | X | 6 |
| Opinion |  |  | X |  |  |  |  | X | 2 |
| Abandoned/Uninterpretable | X |  | X |  | X |  |  | X | 4 |
| Yes-No Question | X | X | X | X |  | X | X | X | 7 |
| Yes answers | X | X | X | X |  | X |  | X | 6 |
| Con. Closing | X | X | X | X |  | X |  | X | 6 |
| Expressive/emotion |  | X |  | X |  | X | X |  | 4 |
| Open question | X | X | X | X | X |  |  | X | 6 |
| Reject | X |  | X |  |  | X |  |  | 3 |
| Apology | X |  | X |  |  |  |  |  | 2 |
| Thanking | X | X |  | X |  |  |  |  | 3 |
| Con. Opening | X | X | X | X |  |  |  |  | 4 |
| Other | X |  | X |  |  | X |  | X | 4 |
| No answer | X | X | X | X |  | X |  | X | 5 |
| Request |  | X |  | X |  |  |  |  | 2 |
| Command / Order |  |  |  |  |  |  | X |  | 1 |
| Self talk | X |  | X |  |  |  |  |  | 2 |
| Backchannel question | X |  | X |  |  |  |  | X | 2 |
| Downplayer |  | X | X | X |  |  |  |  | 3 |
| Accept |  |  | X |  |  | X |  | X | 3 |
| Appreciation |  |  | X |  |  |  |  | X | 2 |
| Non verbal |  |  | X |  |  |  |  | X | 2 |
| wh-question |  |  | X |  |  | X | X | X | 4 |
| Greet |  |  |  |  |  | X |  |  | 1 |
| Emphasis |  |  |  |  |  | X |  |  | 1 |
| Clarify |  |  |  |  |  | X |  |  | 1 |
| Hedge | X |  | X |  |  |  |  | X | 3 |
| Continuer |  |  |  |  |  | X |  | X | 2 |

**Table 3: DA tags analysis of mostly cited research papers**

In the first iteration there were 16 dialogue act tags in the tag set filtered out from the above analysis (table 3). After tagging the sample corpus of 500 utterances the results were given in the table 4 (A graphical representation: Figure 4).

|  |  |  |
| --- | --- | --- |
| Tag | Percentage | Occurrences |
| statement | 42.8 | 214 |
| request | 21 | 105 |
| abandoned/uninterpretable | 13 | 65 |
| open question | 4.8 | 24 |
| yes-no question | 4.2 | 21 |
| Back-channel/acknowledge | 3.4 | 17 |
| command/order | 2 | 10 |
| opinion | 2 | 10 |
| thanking | 1.2 | 6 |
| no answer | 1.2 | 6 |
| expressive | 1.2 | 6 |
| yes answers | 1 | 5 |
| conventional closing | 0.8 | 4 |
| reject | 0.6 | 3 |
| apology | 0.4 | 2 |
| conventional opening | 0.4 | 2 |
|  | **100%** | **500** |
| kappa value | **0.612586** |  |

**Table 4: Tag Set Selection Iteration 01**



**Figure 4: Tag Set Selection Iteration 01**

After the analysis of results it was observed that Request and Command/Order tags should be merged into a single tag because although for English there is a clear separation in utterances between these two tags. Most of the Requests include the word “Please” or a similar phrase in contrast to Command/Orders where it does not. In Sinhala, different forms of the same word is used to indicate whether it is a request or a command. For example, වහන්න (wahanna) is used in requests in a polite manner to say close something (a door) where වහපං (wahapan) is used in orders. And as for the sake of completeness in the second iteration two new dialogue act tags were introduced and tested the tag set again with a different set of dialogue act tags.

It should also be noted here that English-Sinhala translation in baiscope.lk is not just a mere one-to-one mapping from English to Sinhala. This is because the translation process is subjective. The translators generate subtitles while watching the movie. Therefore they capture the prosodic information in the Sinhala subtitles to a great extent. For example, consider a movie scene where an actor asks another actor to "close that door" in a very harsh tone. The corresponding Sinhala subtitle uses command-type words “දොර වහන්න” (dora wahapan) instead of request-type words “දොර වහපං” (dora wahanna).

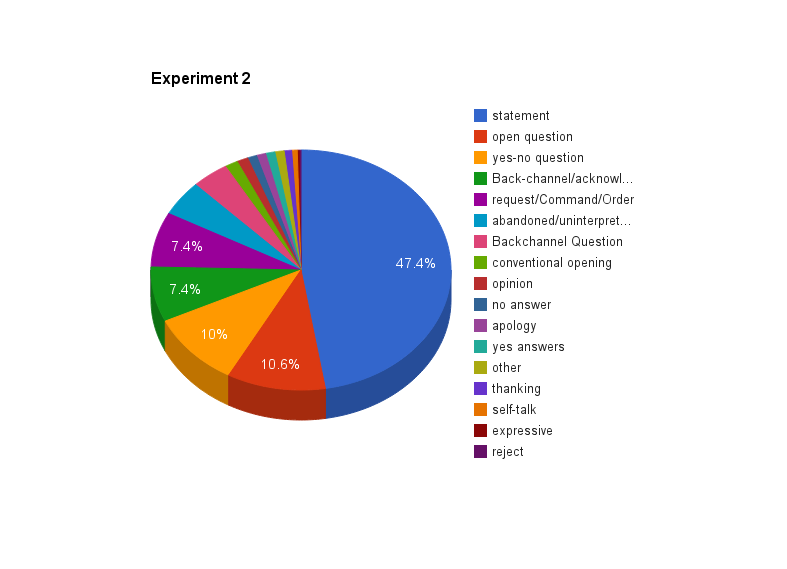
1. Backchannel Question
2. Self-Talk

The rate of occurrence of Backchannel Questions are comparatively high in Sinhala. So we introduced it as a separate tag. Backchannel Questions are Back-Channels or Acknowledges in question form. For example in Sinhala conversations we often come across the phrase “එහෙමද?” (ehemada?) in response, roughly it means “is it?”.

The results of the second iteration after tagging the different sample corpus of 500 utterances the results were given in the table 5 (A graphical representation: Figure 5).

|  |  |  |
| --- | --- | --- |
| **Tag** | **Percentage** | **Count** |
| statement | 47.4 | 237 |
| open question | 10.6 | 53 |
| yes-no question | 10 | 50 |
| Back-channel/acknowledge | 7.4 | 37 |
| request/Command/Order | 7.4 | 37 |
| abandoned/uninterpretable | 4.8 | 24 |
| Backchannel Question | 4 | 20 |
| conventional opening | 1.4 | 7 |
| opinion | 1.2 | 6 |
| no answer | 1 | 5 |
| apology | 1 | 5 |
| yes answers | 1 | 5 |
| other | 1 | 5 |
| thanking | 0.8 | 4 |
| self-talk | 0.6 | 3 |
| expressive | 0.2 | 1 |
| reject | 0.2 | 1 |
|  | 100 | 500 |
| **Kappa Value** | 0.685451549 |  |

**Table 5: Tag Set Selection Iteration 02**



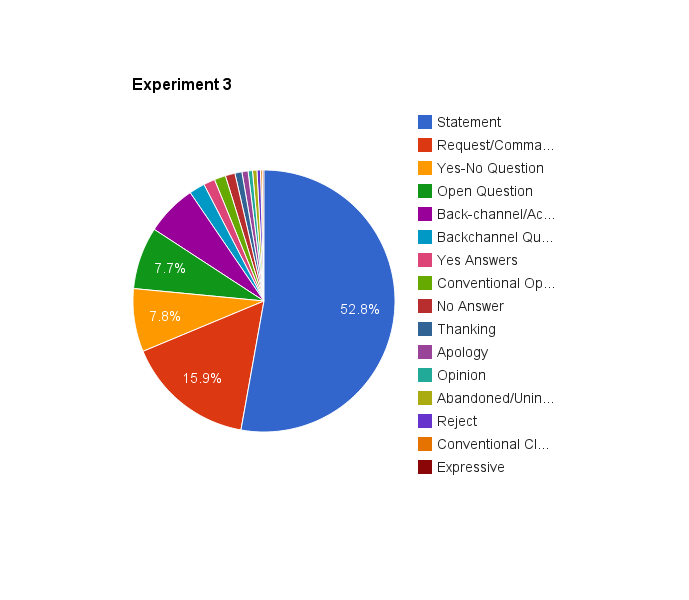
**Figure 5: Tag Set Selection Iteration 02**

After the analysis of results of iteration two that was observed that the two dialogue act tags ‘Other’ and ‘Abandoned/Uninterpretable’ should be merged into a single tag because of their rare occurrence. Although it was observed in iteration one that the self-talk tag was required, in this iteration it was observed that for the general case it was not very frequent. So it has also been removed and conducted the third iteration.

The third iteration was also done using a different sample of the corpus of 500 utterances. The results after tagging are given in the table 6 (A graphical representation: Figure 6).

|  |  |  |
| --- | --- | --- |
| **Tag** | **Percentage** | **Count** |
| Statement | 52.8139 | 488 |
| Request/Command/Order | 15.9091 | 147 |
| Yes-No Question | 7.7922 | 72 |
| Open Question | 7.684 | 71 |
| Back-channel/Acknowledge | 6.2771 | 58 |
| Backchannel Question | 1.9481 | 18 |
| Yes Answers | 1.4069 | 13 |
| Conventional Opening | 1.4069 | 13 |
| No Answer | 1.1905 | 11 |
| Thanking | 0.8658 | 8 |
| Apology | 0.7576 | 7 |
| Opinion | 0.5411 | 5 |
| Abandoned/Uninterpretable/Other | 0.5411 | 5 |
| Reject | 0.4329 | 4 |
| Conventional Closing | 0.2165 | 2 |
| Expressive | 0.2165 | 2 |
|  | **100%** | **924** |
| **Kappa Value** | **0.768569** |  |

**Table 6: Tag Set Selection Iteration 03**

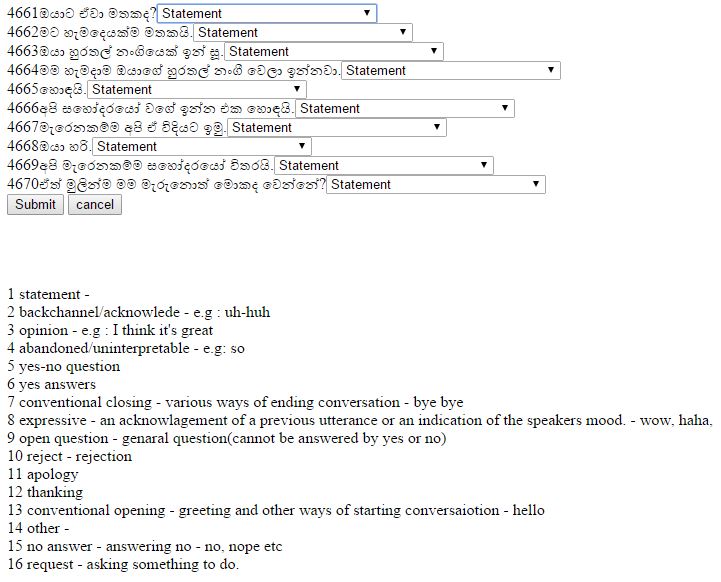
****

**Figure 6: Tag Set Selection Iteration 03**

At the end of this iteration it was observed that this dialogue act tag set is comprehensive and it is suitable to tag the Sanwada corpus for the research work.

To make it easier for the volunteer annotators to tag the corpus, a simple web based tagging tool was implemented. There the tool provides separate logins to the system for the separate annotator. Two sample views of the tagging tool is provided in the figure 7 and figure 8.

After log into the tool the annotator will be provided with 10 utterances in the order they are presented in the corpus. Then the annotator will be able to select the suitable dialogue act tag from the dropdown menu and submit the utterances to the system. Then the system will provide the next 10 utterances to the annotator.



**Figure 7: Tagging tool inner view**

This tagging tool was also included the functionality of computing the inter annotator agreement (the Flies kappa) value. After all four annotator are done with their task the tool will compute and tag each utterance using the most voted by the annotators. That is if all four annotator have selected the same tag the utterance will be tagged by the same tag. If three annotator have used some tag and the other one selected a different tag the utterance will be tagged using the tag selected by the majority. Same method will be applied for scenario of two annotators selecting one tag and the other two selecting two different tags. If all of them selected four different tags for the same utterance or if two of them selected one tag and the other two selected a different tag then those utterances will be presented at the end for the manual tagging.



**Figure 8: Tagging tool basic functionalities**

After tagging the complete corpus manually, we have calculated the inter-annotator agreement among them using Fleiss kappa [44] value and the agreement was 0.8161. To calculate the kappa value we implemented a tool based on the equations introduced by Fleiss [44].

## Feature Selection

The section **2.5** on **Feature Selection** discusses what are features, how can they be used for the dialogue acts classification, the different types of features and the features used for the related studies for English language.

### Identified features from related work for English

There have been identified 14 features that can be used in textual dialogue act recognition from previous studies [39] [45]. Among those 14 features we selected only 7 features for our study considering the applicability to Sinhala and other few concerns that are discussed below. Table 7 lists these features along with their selection status.

|  |  |
| --- | --- |
| **Feature** | **Status** |
| Number of words in the segment | Selected |
| Bigrams/Trigrams of words | Selected |
| Previous Dialogue Act | Selected |
| Verb of the Sentence | Selected |
| Punctuation marks | Selected |
| Grammar pattern | Selected |
| Frequent words for each tag | Selected |
| First two words | Not-selected |
| Last two words | Not-selected |
| First verb type/ Second verb type | Not-selected |
| Words in last 10 Dialogue Acts | Not-selected |
| N-grams of previous Dialogue Acts | Not-selected |
| Bag-of –words | Not-selected |
| Unigrams | Not-selected |

**Table 7: Selected Features**

In Table 7, since we are using the Bigrams as a feature, feature 8 and 9 were omitted. Feature 10 is omitted due to the unavailability of Sinhala PoS tagger. Taking previous Dialogue Acts as features can introduce a cumulative error as described by Lendvai [46]. Unigrams are ineffective for long utterance, although their effectiveness has been shown for chat messages [47].

### Exclusive features for Sinhala

Last letter of the last word of the utterance is one feature that we have identified. Unlike in English, the last letter of the utterance makes a big impact on the dialogue act of the utterance. For instance most of the Yes/No questions ends with the letter ‘ද’(da), most of Request/Command/Order ends with one of the letters ‘න්’(n), ‘න’(na), or ‘ නු’(nu), most of Open questions end with ‘නේ’(ne). Not only the last letter but also the last word of an utterance is an exclusive feature for Sinhala.

The presence of specific Sinhala cue phrases is another identified feature. Table 8 lists some identified cue phrase sets.

|  |  |  |
| --- | --- | --- |
| **Sinhala cue phrase(s)** | **Phonetic Pronunciation** | **English cue phrase** |
| **ඇත්තෙන්ම** | aeththenma | actually |
| **සහ, හා** | saha, haa | and |
| **නිසා, හින්ද** | nisa, hinda | because |
| **එසේම** | esema | also |
| **එහෙත්, නමුත්** | eheth, namuth | but |
| **වගේ, වැනි, වාගේ** | wage, waeni, waage | like |
| **ඉතින්, එවිට** | ithin, ewita | then |
| **හෝ** | ho | or |
| **හරි** | hari | well |
| **එනිසා, එබැවින්** | enisaa, ebawin | so |

**Table 8: Cue Phrases**

### Identified features

Next follows all the major features used for dialog act recognition.

1. Cue Phrases: presence of connective expressions.
2. Number of words in the segment:  self-explanatory
3. Bigrams/Trigrams of words: Adjacent two words in an utterance is considered as a bigram, likewise trigram is adjacent three words.
4. Previous Dialogue Act: The dialogue act of the previous utterance
5. Verb of the Sentence: self-explanatory
6. Punctuation marks: The appearance of the question mark, exclamation mark, Full stop, etc. in the utterance. In Sinhala same punctuation marks are used as in English.
7. Grammar pattern: The Sinhala grammar pattern(s) of the sentences in the utterance
8. Last word of the utterance: self-explanatory
9. Frequent words for each tag: For each tag the most frequent words appear in the training set of utterance.
10. End letter of the last word of the sentence: self-explanatory.

### Feature set selection

The idea of the experiment is to identify the most contributing features for classifying and the most effective combinations of the features. From the aforementioned 10 features, 8 were selected based on the performance evaluation. Because with 10 features, it is computationally expensive than for 8 features to go through all possible combinations.[[5]](#footnote-5)

We used WEKA [48] Java library for classification. To achieve above described task we used InfoGain Attribute Evaluator of WEKA and obtained the InfoGain values. Table 9 displays the results. The InfoGain value evaluates the worth of a feature by measuring the information gain resulted only by that particular feature. For example, a feature with an InfoGain value of 1 means that all of the information available in that feature contributes to classification, though it does not mean that the use of that feature alone is able to conduct the entire classification.

|  |  |  |
| --- | --- | --- |
| **Rank** | **Feature** | **InfoGain** |
| 1 | Punctuation marks | 0.71 |
| 2 | Last word of the utterance | 0.60 |
| 3 | Frequent words for each tag | 0.42 |
| 4 | Trigrams/Bigrams | 0.31 |
| 5 | Last letter of the last word of the sentence | 0.30 |
| 6 | Verb of the Sentence | 0.24 |
| 7 | Number of words in the segment | 0.18 |
| 8 | Cue Phrases | 0.17 |

**Table 9: Individual Feature Performance**

From this result set we can observe that the most contributing feature for the task is Punctuation marks. In the subtitles that we used has been properly written with the use of punctuation marks. This particular feature has been effective in distinguishing questions (Open Question, Yes/No Questions and Back-channel Questions) from other tags. Some of the features that we identified as exclusive features for Sinhala (last word of the utterance and last letter of the utterance) also contributes a considerable amount.

Frequent words for each tag feature keeps track of the most frequent words used in the entire corpus and uses the presence of those words in a particular utterance as a feature for the classifier. For this task we used WEKA’s StringToWordVector option with the word count of 100. This feature has not been widely used in related work but we can observe that this feature works well.

At that time there were limitation on finding the Verb of the sentence precisely such as lack of resources for PoS tagging for Sinhala. Therefore we used a set of commonly used Sinhala verbs to check the presence of those verbs in a given utterance as feature. (In the Symposium on Language Technology South Asia [January 02, 2015] Mr. A.J.P. Monoj Prasad Jayaweera and Prof. N.G.J. Dias did present their work on statistical machine learning using HMM for part of speech tagging.)

From the above mentioned features we have selected the best performing six features listed in the Table 10 by testing the all combinations of features on a selected classifier.

|  |
| --- |
| **Feature** |
| Punctuation marks |
| Last word of the utterance  Trigrams/Bigrams  Last letter of the last word of the sentence  Frequent words for each tag |
| Cue Phrases |

**Table 10: Best Performing Features**

We have used the J48 WEKA classifier to perform this test. For the 8 different features there are 256 different combinations of feature sets. We went through all these different combinations and classified them using a trained J48 classifier. The feature mentioned in the Table 6 yielded the maximum accuracy on the testing set. This feature set achieved F-measure value of 0.755 with a precision 0.788 and recall 0.755.

# Results

For classification WEKA uses various classifiers according to the methods discussed in Section **2.7 Classifier Selection**. For selecting the best classifier for the best feature set we iterated through all the feature combinations through different classifier measuring the accuracy, F-measure and few other measurements. Following are part (only top ten feature combinations for each classifier) of the complete result set that was gained in this evaluation.

Classifier : weka.classifiers.trees.J48

78.92% F->0.755 R->0.789 P->0.773 [segmentLength, punctuation, verb, bagOfWords]

78.82% F->0.755 R->0.788 P->0.775 [segmentLength, lastWord, punctuation, verb, ngrams, bagOfWords]

78.82% F->0.753 R->0.788 P->0.771 [segmentLength, punctuation, verb, ngrams, bagOfWords]

78.71% F->0.757 R->0.787 P->0.762 [segmentLength, punctuation, lastletter, verb, bagOfWords]

78.61% F->0.756 R->0.786 P->0.761 [segmentLength, punctuation, lastletter, verb, ngrams, bagOfWords]

78.5% F->0.751 R->0.785 P->0.769 [segmentLength, punctuation, cuephrases, verb, bagOfWords]

78.4% F->0.75 R->0.784 P->0.768 [segmentLength, punctuation, cuephrases, verb, ngrams, bagOfWords]

78.4% F->0.75 R->0.784 P->0.777 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams, bagOfWords]

78.4% F->0.75 R->0.784 P->0.758 [segmentLength, lastWord, punctuation, verb, bagOfWords]

78.19% F->0.752 R->0.782 P->0.757 [segmentLength, punctuation, lastletter, cuephrases, verb, ngrams, bagOfWords]

Classifier : weka.classifiers.trees.RandomForest

79.23% F->0.776 R->0.792 P->0.78 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

78.71% F->0.768 R->0.787 P->0.769 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb]

78.09% F->0.766 R->0.781 P->0.763 [segmentLength, lastWord, punctuation, lastletter, verb]

77.99% F->0.755 R->0.78 P->0.758 [segmentLength, punctuation, lastletter, cuephrases, verb]

77.57% F->0.751 R->0.776 P->0.753 [segmentLength, punctuation, lastletter, verb]

77.47% F->0.758 R->0.775 P->0.757 [segmentLength, lastWord, punctuation, lastletter, cuephrases]

77.47% F->0.756 R->0.775 P->0.752 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

77.15% F->0.754 R->0.772 P->0.754 [segmentLength, lastWord, punctuation, lastletter, cuephrases, ngrams]

77.15% F->0.744 R->0.772 P->0.747 [segmentLength, punctuation, verb, ngrams]

77.05% F->0.751 R->0.771 P->0.746 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

Classifier : weka.classifiers.rules.PART

78.61% F->0.756 R->0.786 P->0.757 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

78.09% F->0.751 R->0.781 P->0.749 [segmentLength, punctuation, lastletter, verb]

77.88% F->0.747 R->0.779 P->0.741 [segmentLength, punctuation, verb, ngrams]

77.67% F->0.751 R->0.777 P->0.753 [segmentLength, punctuation, lastletter, verb, ngrams]

77.67% F->0.758 R->0.777 P->0.761 [segmentLength, lastWord, punctuation, lastletter, verb]

77.67% F->0.757 R->0.777 P->0.753 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb]

77.36% F->0.753 R->0.774 P->0.747 [segmentLength, lastWord, punctuation, verb, ngrams]

77.05% F->0.74 R->0.771 P->0.749 [segmentLength, punctuation, ngrams]

77.05% F->0.744 R->0.771 P->0.759 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

76.64% F->0.746 R->0.766 P->0.746 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

Classifier : weka.classifiers.trees.LMT

79.44% F->0.762 R->0.794 P->0.78 [segmentLength, punctuation, cuephrases, verb, ngrams]

79.44% F->0.765 R->0.794 P->0.772 [segmentLength, punctuation, lastletter, verb, ngrams]

79.44% F->0.763 R->0.794 P->0.774 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

79.44% F->0.765 R->0.794 P->0.772 [segmentLength, punctuation, lastletter, cuephrases, verb, ngrams]

79.34% F->0.757 R->0.793 P->0.775 [segmentLength, punctuation, cuephrases, verb]

79.34% F->0.757 R->0.793 P->0.772 [segmentLength, punctuation, lastletter, verb]

79.34% F->0.757 R->0.793 P->0.775 [segmentLength, punctuation, verb]

79.34% F->0.763 R->0.793 P->0.77 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

79.34% F->0.758 R->0.793 P->0.769 [segmentLength, punctuation, verb, ngrams]

79.23% F->0.758 R->0.792 P->0.769 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

Classifier : weka.classifiers.rules.DecisionTable

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, cuephrases, verb]

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, verb]

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, verb, ngrams]

75.39% F->0.721 R->0.754 P->0.756 [segmentLength, lastWord, punctuation, lastletter, verb]

75.39% F->0.721 R->0.754 P->0.757 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb]

75.39% F->0.721 R->0.754 P->0.756 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

75.39% F->0.721 R->0.754 P->0.757 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

75.39% F->0.721 R->0.754 P->0.757 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

75.39% F->0.721 R->0.754 P->0.756 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

Classifier : weka.classifiers.functions.SMO

76.95% F->0.708 R->0.769 P->0.728 [segmentLength, punctuation, cuephrases, verb, ngrams]

76.84% F->0.706 R->0.768 P->0.714 [segmentLength, punctuation, verb, ngrams]

76.74% F->0.707 R->0.767 P->0.711 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

76.74% F->0.707 R->0.767 P->0.723 [segmentLength, lastWord, punctuation, verb, ngrams]

76.64% F->0.708 R->0.766 P->0.718 [segmentLength, lastWord, punctuation, verb]

76.64% F->0.708 R->0.766 P->0.718 [segmentLength, punctuation, verb]

76.64% F->0.706 R->0.766 P->0.718 [segmentLength, lastWord, punctuation, cuephrases, verb]

76.32% F->0.699 R->0.763 P->0.705 [segmentLength, punctuation, cuephrases, verb]

76.12% F->0.703 R->0.761 P->0.723 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

76.12% F->0.702 R->0.761 P->0.71 [segmentLength, punctuation, lastletter, verb, ngrams]

Classifier : weka.classifiers.trees.DecisionStump

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation, verb]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation, cuephrases]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation, lastletter, cuephrases, verb]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, lastWord, punctuation, cuephrases, ngrams]

63.86% F->0.501 R->0.639 P->0.413 [lastWord, punctuation, lastletter, cuephrases]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation, lastletter, cuephrases, verb, ngrams]

63.86% F->0.501 R->0.639 P->0.413 [lastWord, punctuation, lastletter, verb]

63.86% F->0.501 R->0.639 P->0.413 [punctuation, verb]

Classifier : weka.classifiers.bayes.NaiveBayes

75.6% F->0.732 R->0.756 P->0.728 [segmentLength, lastWord, punctuation, lastletter, ngrams]

75.39% F->0.731 R->0.754 P->0.728 [segmentLength, punctuation, lastletter, ngrams]

75.29% F->0.726 R->0.753 P->0.722 [segmentLength, lastWord, punctuation, ngrams]

75.18% F->0.726 R->0.752 P->0.72 [segmentLength, punctuation, ngrams]

74.97% F->0.732 R->0.75 P->0.73 [segmentLength, punctuation, lastletter, verb, ngrams]

74.87% F->0.731 R->0.749 P->0.734 [segmentLength, lastWord, punctuation, lastletter]

74.56% F->0.725 R->0.746 P->0.727 [segmentLength, lastWord, punctuation]

74.45% F->0.725 R->0.745 P->0.72 [segmentLength, punctuation, verb, ngrams]

74.45% F->0.729 R->0.745 P->0.732 [segmentLength, punctuation, lastletter]

74.35% F->0.732 R->0.744 P->0.735 [segmentLength, punctuation, lastletter, verb]

Classifier : weka.classifiers.trees.HoeffdingTree

67.71% F->0.577 R->0.677 P->0.522 [punctuation, lastletter]

66.98% F->0.58 R->0.67 P->0.537 [segmentLength, lastWord, punctuation, lastletter]

65.42% F->0.567 R->0.654 P->0.514 [lastWord, punctuation, lastletter]

64.38% F->0.532 R->0.644 P->0.503 [lastWord, punctuation]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation, verb]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation, cuephrases]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, punctuation, lastletter, cuephrases, verb]

63.86% F->0.501 R->0.639 P->0.413 [segmentLength, lastWord, punctuation, cuephrases, ngrams]

Classifier : weka.classifiers.trees.REPTree

78.19% F->0.731 R->0.782 P->0.761 [segmentLength, punctuation, verb, ngrams]

78.19% F->0.731 R->0.782 P->0.761 [segmentLength, punctuation, cuephrases, verb, ngrams]

78.09% F->0.744 R->0.781 P->0.762 [segmentLength, lastWord, punctuation, cuephrases]

78.09% F->0.744 R->0.781 P->0.762 [segmentLength, lastWord, punctuation]

77.99% F->0.739 R->0.78 P->0.758 [segmentLength, lastWord, punctuation, cuephrases, ngrams]

77.99% F->0.739 R->0.78 P->0.758 [segmentLength, lastWord, punctuation, ngrams]

77.57% F->0.725 R->0.776 P->0.754 [segmentLength, punctuation, cuephrases, verb]

77.57% F->0.725 R->0.776 P->0.754 [segmentLength, punctuation, verb]

77.47% F->0.733 R->0.775 P->0.76 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

77.47% F->0.733 R->0.775 P->0.76 [segmentLength, lastWord, punctuation, verb, ngrams]

Classifier : weka.classifiers.rules.DecisionTable

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, verb]

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, verb, ngrams]

76.12% F->0.727 R->0.761 P->0.757 [segmentLength, lastWord, punctuation, cuephrases, verb]

75.39% F->0.721 R->0.754 P->0.756 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

75.39% F->0.721 R->0.754 P->0.756 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb]

75.39% F->0.721 R->0.754 P->0.757 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

75.39% F->0.721 R->0.754 P->0.757 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb]

75.39% F->0.721 R->0.754 P->0.756 [segmentLength, lastWord, punctuation, lastletter, verb]

75.39% F->0.721 R->0.754 P->0.756 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

Classifier : weka.classifiers.functions.SimpleLogistic

79.44% F->0.765 R->0.794 P->0.772 [segmentLength, punctuation, lastletter, cuephrases, verb, ngrams]

79.44% F->0.765 R->0.794 P->0.772 [segmentLength, punctuation, lastletter, verb, ngrams]

79.44% F->0.762 R->0.794 P->0.78 [segmentLength, punctuation, cuephrases, verb, ngrams]

79.44% F->0.763 R->0.794 P->0.774 [segmentLength, lastWord, punctuation, lastletter, cuephrases, verb, ngrams]

79.34% F->0.763 R->0.793 P->0.77 [segmentLength, lastWord, punctuation, lastletter, verb, ngrams]

79.34% F->0.758 R->0.793 P->0.769 [segmentLength, punctuation, verb, ngrams]

79.34% F->0.757 R->0.793 P->0.772 [segmentLength, punctuation, lastletter, verb]

79.34% F->0.757 R->0.793 P->0.775 [segmentLength, punctuation, cuephrases, verb]

79.34% F->0.757 R->0.793 P->0.775 [segmentLength, punctuation, verb]

79.23% F->0.758 R->0.792 P->0.769 [segmentLength, lastWord, punctuation, cuephrases, verb, ngrams]

Observing the above result set the best performing feature set and the best performing classifier was selected. For classification task we have used 8000 utterances as training set and 4000 utterances as testing set. As the first step we have tested the classification accuracy by just using the features used for dialogue act recognition in English. From the best performing features stated in the Table 6, Punctuation marks, Trigrams/Bigrams and Frequent words for each tag are the three features used in the related work. The other three features are specific for Sinhala. Using those three features used for English we were able to gain an accuracy of 71.14% in classification using the J48 classifier. Then we have used all six features and classified using the same classifier and we were able to improve the accuracy to 78.92%.

As the next step we have used the same feature set and classified the same data set using different classifiers to model the performance of different classifiers on Sinhala.

|  |  |  |  |
| --- | --- | --- | --- |
| **Classifier** | **Recall** | **Precision** | **F-measure** |
| RandomForest | 0.792 | 0.780 | 0.776 |
| SimpleLogistic | 0.794 | 0.772 | 0.765 |
| LMT | 0.794 | 0.760 | 0.762 |
| PART | 0.786 | 0.757 | 0.756 |
| J48 | 0.789 | 0.773 | 0.755 |
| NaiveBayes | 0.756 | 0.728 | 0.732 |
| REPTree | 0.782 | 0.761 | 0.731 |
| DecisionTable | 0.761 | 0.757 | 0.727 |
| SMO | 0.769 | 0.728 | 0.708 |
| DecisionStump | 0.639 | 0.413 | 0.501 |
| HoeffdingTree | 0.677 | 0.522 | 0.577 |

**Table 11: Classifier Performance**

Table 12 lists the classifiers in the descending order of F-measure value. F-measure represents a value of accuracy of the tests performed which is calculated using recall and precision value. We can observe that SimpleLogistic and LMT classifiers gives the highest recall value. That means they have identified more correctly tagged utterances compared to other classifiers. Also, we noticed that decision tree based classifiers performs well for Sinhala Classification as shown in the table.

## Performance Analysis on Classifiers

A performance analysis for all the classifier have been done at the final phase of this study. For that analysis a core-i3 laptop computer with 4 GB RAM was used. The above table 12 represents the results of the analysis including the details of CPU and Memory consumptions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Classifier** | **CPU Usage (%)** | **Memory Consumption (%)** | **Elapsed Time (S)** |
| J48 | 44.83 | 3.55 | 12 |
| RandomForest | 39 | 5.8 | 6 |
| PART | 28.47 | 7.62 | 119 |
| LMT | 26.21 | 11.67 | 651 |
| DecisionTable | 26.25 | 16.13 | 70 |
| SimpleLogistic | 26.05 | 17.15 | 105 |
| SMO | 26 | 17.5 | 4 |
| DecisionStump | 26 | 13.5 | 2 |
| NaiveBayes | 26 | 13.5 | 4 |
| HoeffdingTree | 26 | 13.51 | 14 |
| REPTree | 26 | 13.6 | 4 |

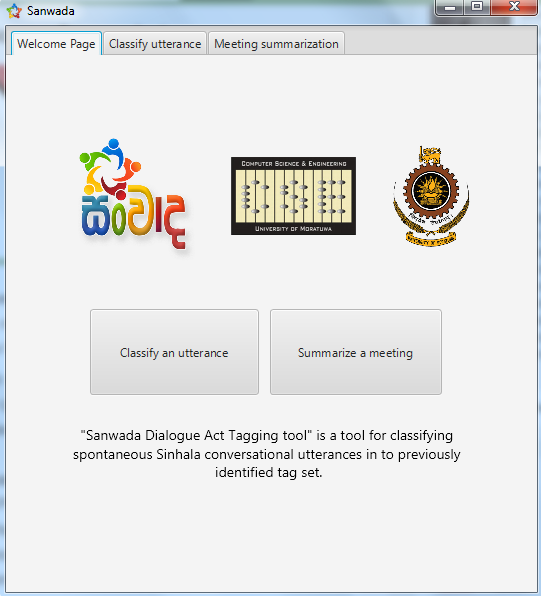
**Table 12: Performance analysis results for classifiers**

The above shown parameters are average value of hundreds of executions. Here you can clearly observed that the support vector machine based and decision tree based classifiers such as SMO, RandomForest and REPTree consuming less average elapsed time compared to other type of classifiers. Average CPU usage of all the classifiers is approximately same except J48 and RandomForest. But these two classifiers have very low average memory consumption.

# **Tools**

The objective of this research is to enable Dialogue Act Recognition in Sinhala. DA recognition can be used in many applications like meeting summarization and intelligent voice assistance. So, we have developed a desktop application and a web application based on our research which can classify a user entered utterance into the appropriate DA type in real time and can summarize meeting according to the user’s wish.

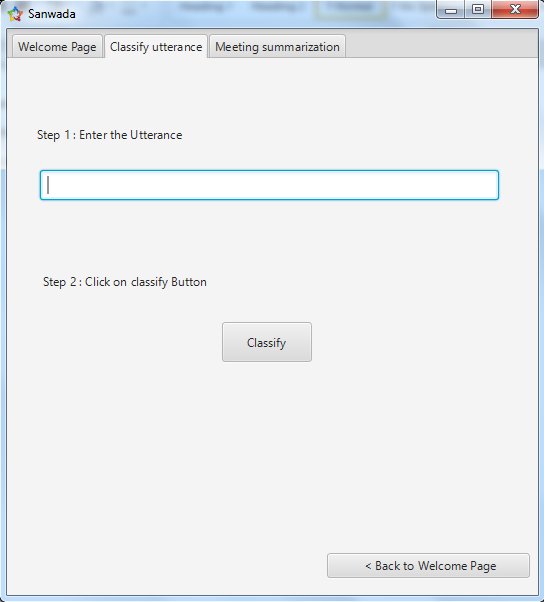
## Desktop Application

This application has two main functionalities such as classify an utterance and meeting summarization. When the user runs the application he will prompted to the following welcome page. The front view of the application is provided in figure 9.

**Figure 9**

### Classify utterance

This function classifies a user entered utterance to appropriate Dialogue Act type in real time. The following is the initial screen of classify utterance functionality. Here user can enter whatever sentence/utterance in Sinhala (figure 10) in the text box and click the classify button to get the DA type.



**Figure 10**

### 

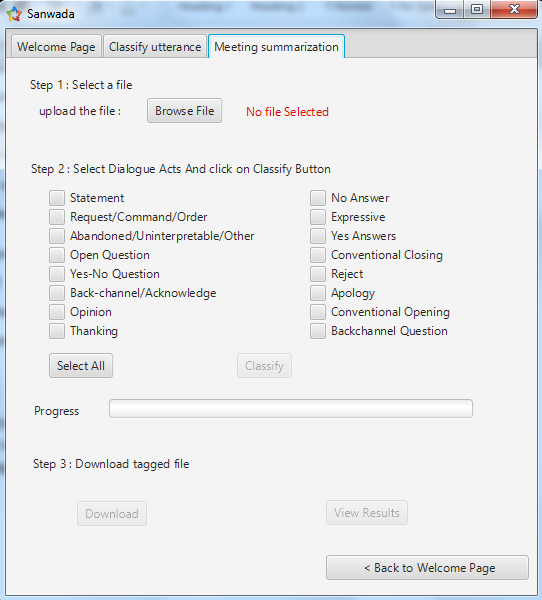
**Figure 11**

Here we have entered a sample utterance “මම ඔහුට කීවෙ නැද්ද?” to show you the process. When the user click the classify button, the application immediately show the appropriate Dialogue Act type in the screen as shown in the above image (figure 11). Here the sample utterance is classified as yes-no question DA type.

### Meeting summarization

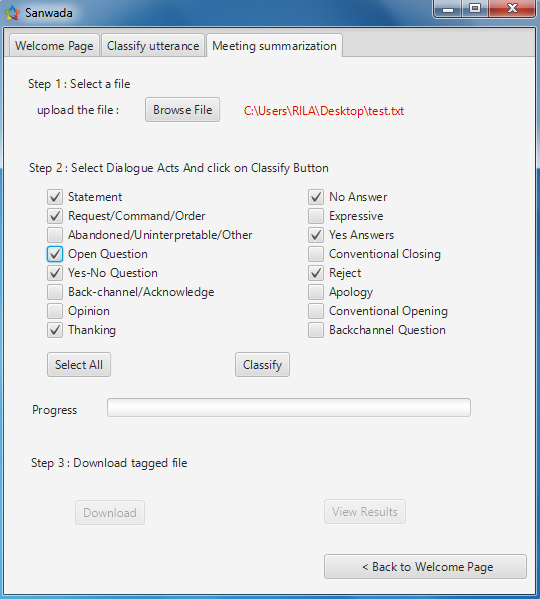
For better or worse, meetings play an integral role in most of our daily lives — they let us share information and collaborate with others to solve a problem, to generate ideas, and to weigh options. Not surprisingly then, there is growing interest in developing automatic methods for meeting summarization. Generating summaries of a particular aspect of a meeting rather than of the meeting as a whole is very important in today’s context. For example, one might want a summary of just the questions asked during the meeting, decisions made during the meeting, and the given answers, the action items that emerged, the ideas discussed, or the hypothesis put forth, etc.

This function does a meeting summarization to a user given text file. User can upload a text file from his computer to the application. This functionality is composed of three steps. The following is the initial screen of the meeting summarization functionality. In the first step, user needs to select a text file which contains the utterances. This shows in red as “No file selected”, if no file is selected. After selecting a file the warning will go away and it will show the path of the file location (figure 12).



**Figure 12**

In the second step, user can tick the dialogue Act types for which he need a summarization. If a summary of all the DA type is needed, user can simply click the “Select All” button which ticks all the available DA types. After pressing the “Classify” button the application process the given text file and list out the utterances under the selected categories of Dialogue Acts (figure 13).

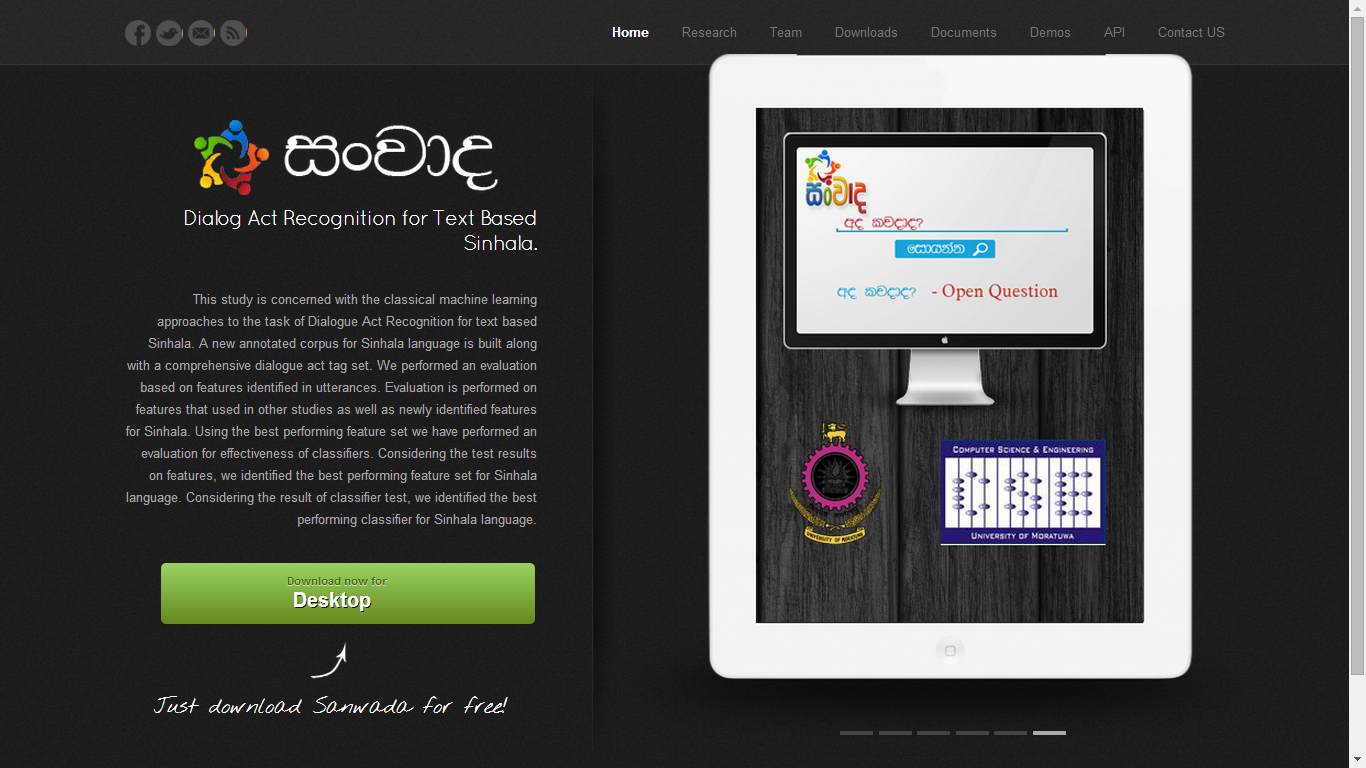


**Figure 13**

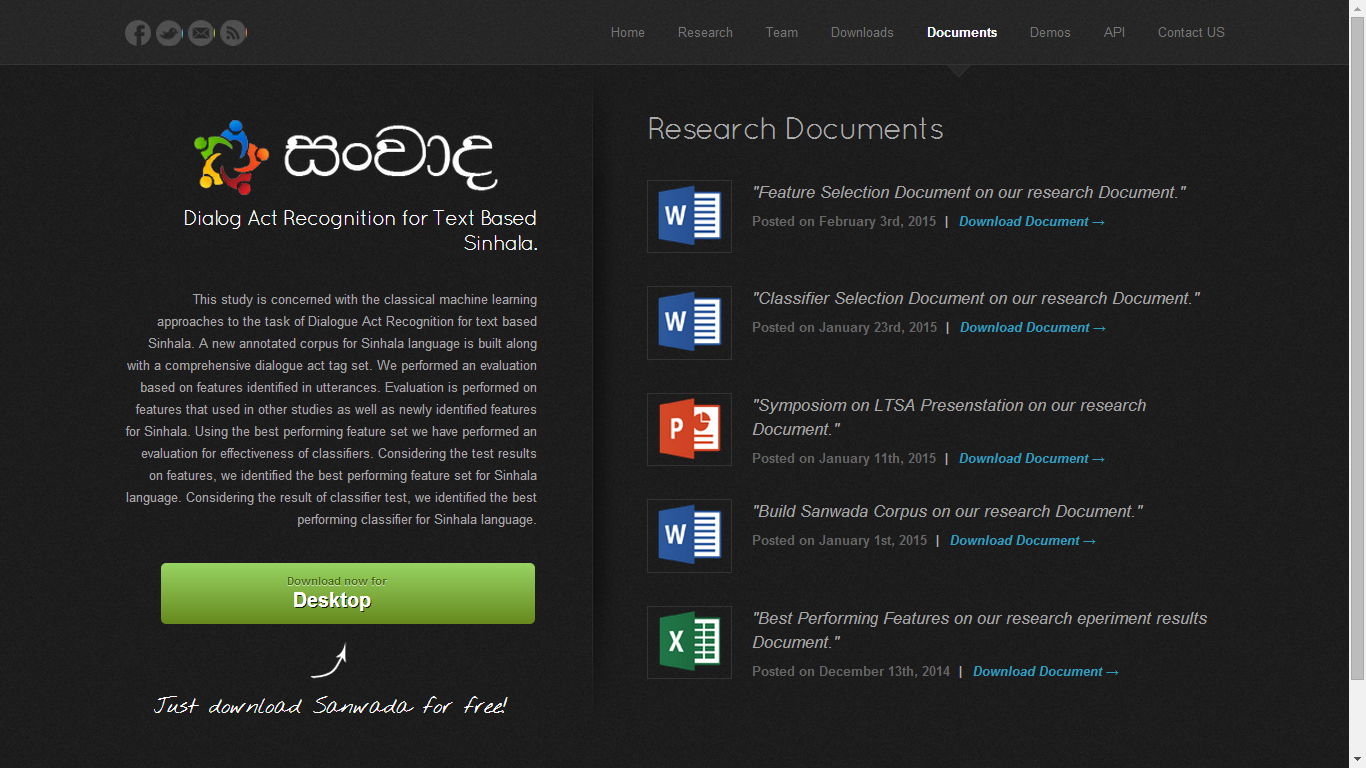
The progress bar displays the progress of classification after clicking the “Classify” button.

In step 3, user can download the classified output or can view the output of classification. These buttons become active only after the progress completed. If the “Download” button is clicked, the application asks to save the output in a user preferred location. If the “View Results” is clicked, the application shows the output of the classification through the application.

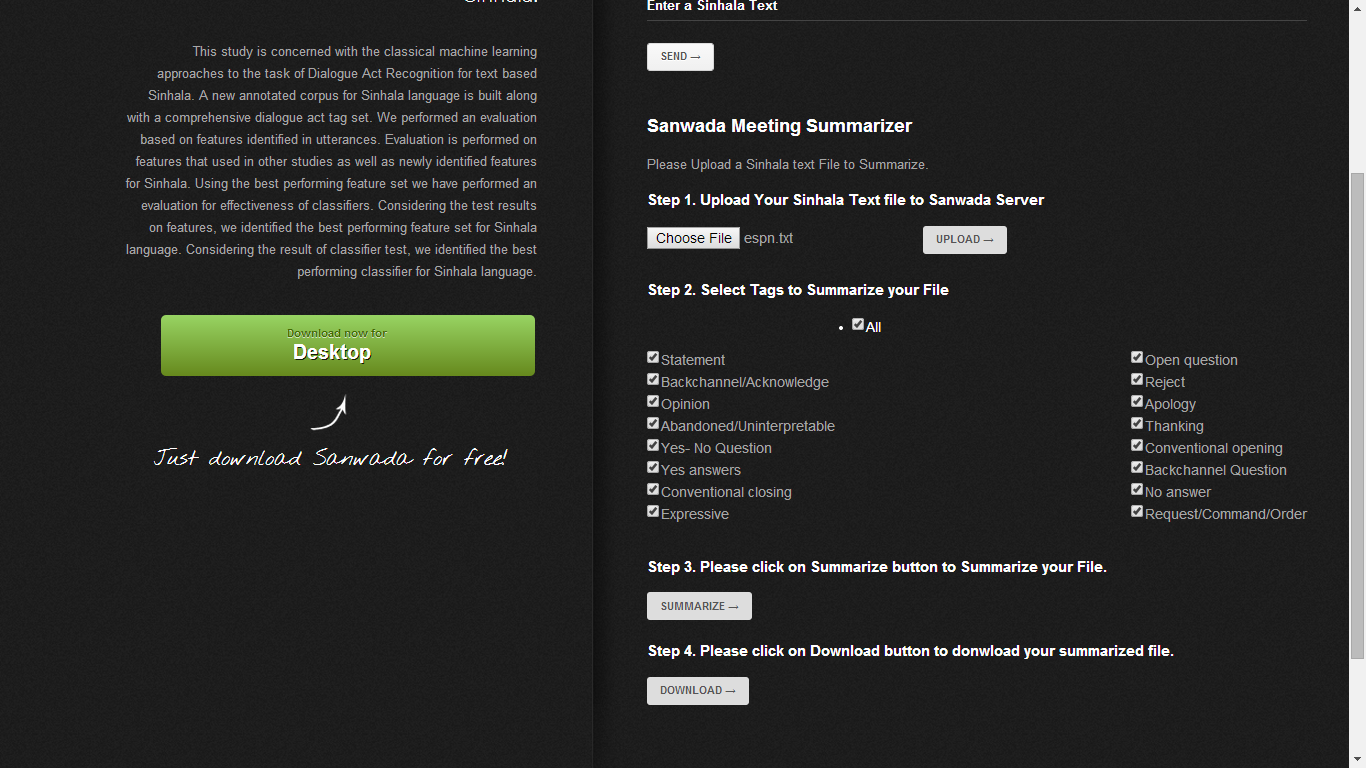
## Web Application and RESTful API

Web application (figure 14, figure 15, figure 16) provides the same functionality as the desktop application. But, it contains articles and documents related to the research work done. So, this is good reference for future researches to do a preliminary study.****

**Figure 14: Home Page**

****

**Figure 15: Important Documents**

****

**Figure 16: Functionality**

# Future Work

In future, we are planning to do following tasks to enrich Dialogue Act recognition accuracy, usefulness and application.

## Taking Prosodic information into the context

Prosodic and acoustic DA classification is based on number of factors such as utterance duration, pitch, pauses, energy, speaking rate and gender that are computed automatically from the speech signal. Prosodic information is vital for DA classification, because word-based classification suffers from recognition errors and some utterances are inherently ambiguous based on words alone. Conversations contain non-linguistic information such as intension, topic change and emphasizing words or phrases, which is mainly expressed by prosody, while written text conveys only linguistic information. For example, some Yes-No-Questions have word sequences identical to those of statements, but can often be distinguished by their prosodic information. Introducing prosodic information into the Dialogue Act recognition process is expected to improve the quality of the classification significantly.

We have achieved an accuracy of 78.68% without prosodic information. In Sinhala, the same word is used in different forms in different situations. With only the prosodic information, we can understand the real expression of the speaker. So, by introducing prosodic information, we can definitely improve the accuracy of the classifier to a greater extent.

## Identify new features

This research has identified 7 general features and 3 Sinhala Specific features. Among this ten, only six features were selected as best performing features and obtained above mentioned accuracy. But, we have planned to look into some exclusive Sinhala features after creating a Sinhala POS tagging tool. Also, Sinhala grammar pattern is complex compared to the English. So, it needs a deeper study of grammar and linguistic patterns to identify some new features. But, exploring these new features will be definitely improving the classifier accuracy by significant value.

## Intelligent voice assistance for Sinhala

Intelligent voice assistance uses a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Web services. The popular application is Siri in ios platform which let your voice to send messages, schedule meetings, place phone calls, and more. This application only supports eight major languages. With the creation of DA recognition, POS tagging tools and some other related researches it’s possible to add Sinhala support for intelligent vice assistances.

## Classifier optimization

This research uses classifiers available in Weka data mining tool to do the classification task. All of these classifiers are general classifier and not optimized for Dialogue Act recognition or Sinhala natural language processing. So, in future we will try to optimize and tune some best performing classifiers for Sinhala Dialogue Act recognition as well as for Sinhala language processing.

# **Conclusion**

In this study we have presented an extensive survey of Dialogue Act Recognition by machine learning approaches and a brief review of related studies, research and techniques. We have explained the importance of recognizing Dialogue Acts in a conversation and the major steps involved with the process of identification. Extensive description about corpora used in various studies for different languages including English, German and French. We have compared the corpora available in the context against utterance count, word count and dialogue type. In this literature we pointed out that there are no corpus available for Sinhala language and we provided the approaches can be used to build a Sinhala corpus.

Selecting features for the classifiers is a very important step in the process of dialogue act recognition using classifiers. We have summarized the feature selection criteria used by the studies done in the context. Then we discussed about different classification techniques can be used and we end the review with a discussion on selecting a classifier and the parameters we can use for measure the classifier performance. We have done an extensive analysis on the performance of features and classifiers. Based on this analysis, we have selected the best performing feature set and classifier.

The outcome of this research is very useful for any Sinhala NLP oriented researcher and we have built tools and an API to ease the DA recognition part. Also, this research ensure the generic features can be applied for its branch and cousin languages, because we have obtained significant amount accuracy by applying only generic features of English. So, researches can apply the generic features of their cousin languages to their NLP researches.

# **References**

[1] Král, P., & Cerisara, C. (2012). Dialogue act recognition approaches. *Computing and Informatics*, *29*(2), 227-250.

[2] Searle, John R. *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press, 1985.

[3] Austin, John Langshaw. *How to do things with words*. Vol. 1955. Oxford university press, 1975.

[4] Allen, J., & Core, M. (1997). Draft of DAMSL: Dialog act mark-up in several layers. *Unpublished manuscript*, *2*.

[5] Jurafsky, D., Shriberg, E., & Biasca, D. (1997). Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. *Institute of Cognitive Science Technical Report*, 97-102.

[6] Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., & Carvey, H. (2004). *The ICSI meeting recorder dialog act (MRDA) corpus*. INTERNATIONAL COMPUTER SCIENCE INST BERKELEY CA.

[7] Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan & Wooters, C. (2003, April). The ICSI meeting corpus. In *Acoustics, Speech, and Signal Processing, 2003. IEEE International Conference on* (Vol. 1, pp. I-364). IEEE.

[8] Klein, A., Maier, E., Maleck, I., Mast, M., & Quantz, J. (1995). *Dialogue acts in VERBMOBIL*. University.

[9] Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992, March). SWITCHBOARD: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on* (Vol. 1, pp. 517-520). IEEE.

[10] Kurematsu, A., Akegami, Y., Burger, S., Jekat, S., Lause, B., MacLaren, & Schultz, T. (2000). VERBMOBIL dialogues: multifaced analysis. In *INTERSPEECH* (pp. 712-715).

[11] Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, & Weinert, R. (1991). The HCRC map task corpus. *Language and speech*, *34* (4), 351-366.

[12] Akker, H., & Schulz, C. (2008). Exploring features and classifiers for dialogue act segmentation. In *Machine Learning for Multimodal Interaction* (pp. 196-207). Springer Berlin Heidelberg.

[13] Poel, M., Stegeman, L., & op den Akker, R. (2007). A support vector machine approach to dutch part-of-speech tagging. In *Advances in intelligent data analysis VII* (pp. 274-283). Springer Berlin Heidelberg.

[14] Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, *19*(2), 313-330.

[15] Jurafsky, D., Shriberg, E., Fox, B., & Curl, T. (1998, August). Lexical, prosodic, and syntactic cues for dialog acts. In *Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers* (pp. 114-120).

[16] Litman, D. J., Hirschberg, J. B., & Swerts, M. (2000, April). Predicting automatic speech recognition performance using prosodic cues. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference* (pp. 218-225). Association for Computational Linguistics.

[17] S. Rosset and L. Lamel, “Automatic detection of dialog acts based on multi-level information,” in icslp, Jeju Island, October 2004, pp. 540–543.

[18] Lendvai, P., van den Bosch, A., & Krahmer, E. (2003). Machine learning for shallow interpretation of user utterances in spoken dialogue systems. In *Proc. of EACL-03 Workshop on Dialogue Systems: interaction, adaptation and styles of management* (pp. 69-78).

[19] Andernach, T. (1996). A machine learning approach to the classification of dialogue utterances.

[20] Zimmermann, M., Hakkani-Tür, D., Shriberg, E., & Stolcke, A. (2006). Text based dialog act classification for multiparty meetings. In *Machine Learning for Multimodal Interaction* (pp. 190-199). Springer Berlin Heidelberg.

[21] Petukhova, Volha Viktarauna. "Multidimensional dialogue modelling." *Promotor: H. Bunt. Tilburg* 1 (2011).

[22] Hirschberg, J., & Litman, D. (1993). Empirical studies on the disambiguation of cue phrases. *Computational linguistics*, *19*(3), 501-530.

[23] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky & Meteer, M. (2000). Dialogue act modelling for automatic tagging and recognition of conversational speech. *Computational linguistics*, *26*(3), 339-373.

[24] Geertzen, J., Petukhova, V., & Bunt, H. (2007, September). A multidimensional approach to utterance segmentation and dialogue act classification. In*Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, Antwerp* (pp. 140-149).

[25] Reithinger, N., & Klesen, M. (1997, September). Dialogue act classification using language models. In *EuroSpeech*.

[26] Samuel, K., Carberry, S., & Vijay-Shanker, K. (1998, August). Dialogue act tagging with transformation-based learning. In *Proceedings of the 17th international conference on Computational linguistics-Volume 2* (pp. 1150-1156). Association for Computational Linguistics.

[27] Keizer, S. (2003). *Reasoning under uncertainty in natural language dialogue using Bayesian networks*. Twente University Press.

[28] Lendvai, P., Bosch, A. V. D., Krahmer, E., & Canisius, S. (2004). Memory-based robust interpretation of recognised speech. In *9th Conference Speech and Computer*.

[29] Fishel, M. (2006). Dialogue Act Recognition Techniques. *GSLT/NGSLT course on dialogue systems: Linkoping University, Sweden*.

[30] Kim, M. J., Rim, H. C., Park, S. Y., & Han, K. S. (2013). N-gram based Dialogue Act Classification Model with Position Numbering. *Advances in Information Sciences & Service Sciences*, *5*(10).

[31] Manning, C. D. (1999). *Foundations of statistical natural language processing*. H. Schütze (Ed.). MIT press.

[32] Rabiner, L., & Juang, B. H. (1986). An introduction to hidden Markov models.*ASSP Magazine, IEEE*, *3*(1), 4-16.

[33] Powers, David Martin. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." (2011).

[34] Cohen (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960:37-46.

[35] Fleiss, J. L. (1981). Statistical methods tor rates and proportions. *Nueva York: Wiley*, *8*.

[36] Fišel, M. (2007). Machine learning techniques in dialogue act recognition. *Eesti* Rakenduslingvistika Ühingu aastaraamat, (3), 117-134.

[37] Balkin, S. D., & Ord, J. K. (2000). Automatic neural network modelling for univariate time series. International Journal of Forecasting, 16(4), 509-515.

[38] Wasala, Ruwan Asanka, et al. "An Open-Source Data Driven Spell Checker for Sinhala." International Journal on Advances in ICT for Emerging Regions (ICTer) 3.1 (2011): 11-24.

[39] Verbree, Daan, Rutger Rienks, and Dirk Heylen. "Dialogue-act tagging using smart feature selection; results on multiple corpora." Spoken Language Technology Workshop, 2006. IEEE. IEEE, 2006.

[40] Herath, Dulip Lakmal, and A. R. Weerasinghe. "A Stochastic Part of Speech Tagger for Sinhala." Proceedings of the 06th International Information Technology Conference (IITC’04), Colombo, Sri Lanka. 2004.

[41] Paññāsāra, O. and Āracci, V. (2011). Siṃhala bhāṣā vikāsaya saha śilālēkana vimarśana. [Koḷamba: Okkampīṭiyē Paññāsāra Himi].

[42] Holman, Eric W., et al. "Automated dating of the world’s language families based on lexical similarity." Current Anthropology 52.6 (2011): 841-875.

[43] Wijesiri, Indeewari, et al. "Building a WordNet for Sinhala."

[44] Fleiss, Joseph L., Bruce Levin, and Myunghee Cho Paik. "The measurement of interrater agreement." Statistical methods for rates and proportions 2 (1981): 212-236.

[45] Rosset, Sophie, and Lori Lamel. "Automatic Detection of Dialog Acts Based on Multi-level Information Ý." (2004).

[46] Lendvai, Piroska, Antal van den Bosch, and Emiel Krahmer. "Machine learning for shallow interpretation of user utterances in spoken dialogue systems." InProc. of EACL-03 Workshop on Dialogue Systems: interaction, adaptation and styles of management, pp. 69-78. 2003.

[47] Ivanovic, Edward. "Automatic instant messaging dialogue using statistical models and dialogue acts." University of Melbourne, Department of Computer Science and Software Engineering, 2008.

[48] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

1. Population aged 10 years and over [↑](#footnote-ref-1)
2. Used google translator [↑](#footnote-ref-2)
3. http://www.baiscopelk.com/ [↑](#footnote-ref-3)
4. http://web.sanwada.robotics.lk/download.html [↑](#footnote-ref-4)
5. For 10 features have to go through 210 i.e. 1024 combinations where for 8 features it’s only 28 i.e. 256 [↑](#footnote-ref-5)